PROPERTIES OF BAHADUR EFFICIENCY

By

RAMON C. LITTELL

Bachelor of Science
Kansas State Teachers College
Emporia, Kansas
1964

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1966

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1970

PROPERTIES OF BAHADUR EFFICIENCY

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGMENTS

My most sincere gratitude is extended to Dr. J. Leroy Folks, whose interest, guidance, and generosity made this research possible.

I also wish to thank Dr. Robert D. Morrison, Dr. Lyle Broemeling and Dr. Odell Walker for serving on my advisory committee.

Further appreciation goes to the National Science Foundation for a Traineeship in 1966-67, for a Summer Research Assistantship in 1969, and for affording me attendance to "A Conference on Limit Theorems in Statistics" at Florida State University in September, 1969.

Special thanks is due my wife, Sue, who has shown continual patience and encouragement during the course of my graduate study.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Many criteria have been devised for the purpose of comparing test statistics. Some of these criteria are based on the asymptotic behavior of the test statistic; that is, on the limiting properties of the test statistic as the sample size approaches infinity. In particular, R. R. Bahadur (1) has proposed two asymptotic criteria, one based on the rate of convergence to zero of the level of significance attained by the test statistic, and the other based on the rate of convergence of an approximate level of significance. One point to notice is that these criteria do not rely on any reference to power or size of the test statistic, although there is a connection. The present study investigates some of the mathematical properties of these criteria.

Approximate Bahadur Efficiency

Denote by $s = (x_1, x_2, \ldots, \text{ad inf})$ a sequence of realizations of the random variables $X_1, X_2, \ldots,$ whose probability distribution P_θ depends on a parameter θ which belongs to a set Θ, and let H be the null hypothesis $H: \theta \in \Theta_0$, where $\Theta_0 \subset \Theta$. For $n = 1, 2, \ldots,$ let $T_n(s)$ be a real valued statistic which depends only on the first n observations $x_1, \ldots, x_n$. Large values of T_n will be taken as significant for rejecting H; that is, lending incredibility to the truth of H.

Following Bahadur (1), $\{T_n\}$ is called a standard sequence if the following conditions are satisfied:

i) There exists a continuous distribution function F such that $F_n(t) \to F(t)$ as $n \to \infty$, where F_n is the distribution function of T_n when $\theta \varepsilon \Theta_0$.

ii) There exists a real number $a > 0$ such that

$$\log[1-F(t)] = -\frac{1}{2} at^2[1+o(1)] \quad \text{as} \quad t \to \infty$$

iii) There exists a positive function $b(\theta)$ on $\Theta - \Theta_0$ such that

$$\frac{T_n}{\sqrt{n}} \to b(\theta) \quad \text{in probability} \quad [\theta] .$$

Then $c^{(a)}(\theta)$ defined by $c^{(a)}(\theta) = 0$ for $\theta \varepsilon \Theta_0$ and $c^{(a)}(\theta)=ab^2(\theta)$ for $\theta\varepsilon\Theta-\Theta_0$ is called the (approximate) slope of $\{T_n\}$. If $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two standard sequences with slopes $c_1^{(a)}(\theta)$ and $c_2^{(a)}(\theta)$, respectively, then $\phi_{12}^{(a)}(\theta)=\dfrac{c_1^{(a)}(\theta)}{c_2^{(a)}(\theta)}$ is called the (approximate) Bahadur efficiency of $\{T_n^{(1)}\}$ with respect to $\{T_n^{(2)}\}$.

It is seen that $1-F(T_n(s))$ is an approximation to the null probability of obtaining a value of T_n larger than the observed value, $T_n(s)$. That is, with $L(t) = 1-F(t)$, $L(T_n(s))$ approximates the level attained by the statistic T_n when s is observed. Bahadur shows [1] that if $K_n(s)$ is defined by $K_n(s) = -2 \log L(T_n(s))$, then $\frac{1}{n} K_n(s) \to c^{(a)}(\theta)$ in probability $[\theta]$, and hence that

$\dfrac{K_n^{(1)}(s)}{K_n^{(2)}(s)} \to \phi_{12}^{(a)}(\theta)$ in probability $[\theta]$. By considering as desirable test statistics which yield small values of $L(T_n(s))$, it follows that large values of $K_n(s)$ are desirable. Thus $T_n^{(1)}$ is judged superior to $T_n^{(2)}$ if $\phi_{12}^{(a)}(\theta) > 1$. This is one sense in which $\phi_{12}^{(a)}(\theta)$ measures the asymptotic relative efficiency of $\{T_n^{(1)}\}$ with respect to $\{T_n^{(2)}\}$.

The function $\psi_{12}(\theta_0) = \lim_{\theta\to\theta_0} \phi_{12}^{(a)}(\theta)$ is of special interest, partly because it is equal to the Pittman efficiency in most cases (1).

Exact Bahadur Efficiency

With $L_n(t)$ defined by $L_n(t) = 1-F_n(t)$, it is seen that $L_n(T_n(s))$ is the (exact) level of significance attained by T_n when s is observed. In typical cases there is a function $c(\theta)$, called the exact slope of T_n, such that $-\frac{2}{n} \log L_n(T_n(s)) \to c(\theta)$ with probability one $[\theta]$. It follows immediately that

$$\frac{-2 \log L_n^{(1)}(T_n^{(1)}(s))}{-2 \log L_n^{(2)}(T_n^{(2)}(s))} \to \frac{c_1(\theta)}{c_2(\theta)} = \phi_{12}(\theta) ,$$

as in the case of approximate slopes.

For fixed α, $0 < \alpha < 1$, let $N(\alpha,s)$ be the smallest integer m such that $L_n(T_n(s)) < \alpha$ for all $n \geq m$; that is, $N(\alpha,s)$ may be regarded as the smallest sample size necessary to attain a level smaller than α for all larger sample sizes when s is observed. Then $\lim_{\alpha\to 0} \frac{-2 \log \alpha}{N(\alpha,s)} = c(\theta)$ with probability one $[\theta]$, and consequently,

$$\lim_{\alpha\to 0} \frac{N^{(2)}(\alpha,s)}{N^{(1)}(\alpha,s)} = \phi_{12}(\theta) \text{ with probability one } [\theta] .$$

For proofs, see Bahadur (1). This is another sense in which $\phi_{12}(\theta)$ measures the relative efficiency of $\{T_n^{(1)}\}$ with respect to $\{T_n^{(2)}\}$.

For a further discussion of $c^{(a)}(\theta)$ and $c(\theta)$, see (2).

Calculation of Slopes

The actual calculation of approximate slopes is generally easy, in contrast to the relative difficulty of calculating exact slopes. Some techniques have been devised which greatly simplify the calculation of exact slopes for certain classes of statistics, for example (3), (8), and (14). A theorem will now be stated, without proof, which utilizes large deviation theory in calculating exact slopes. For a complete proof, see I. R. Savage (13).

Theorem 1.1: If $\{T_n\}$ is a sequence of statistics which satisfy the following two properties:

i) There exists a function $b(\theta)$, $0 < b(\theta) < \infty$, such that $\frac{T_n}{\sqrt{n}} \to b(\theta)$ with probability one $[\theta]$.

ii) There exists a continuous function $f(t)$ such that for each t in some neighborhood of $b(\theta)$,

$$\lim_{n\to\infty} \frac{-1}{n} \log P_0\{T_n \geq \sqrt{n}\, t\} = f(t)$$

Then the exact slope of $\{T_n\}$ is given by $c(\theta) = 2f(b(\theta))$.

An example is now given to illustrate the use of Theorem 1.1.

Example 1.1: Let Y_i , $i = 1,2,\ldots$, be independent random variables, each distributed exponentially with parameter $\lambda > 0$, that is $P_\lambda\{Y_i \le t\} = 1-e^{-\lambda t}$, and consider the hypotheses $H: \lambda = \lambda_0$ and $A: \lambda < \lambda_0$, where $\lambda_0 > 0$ is fixed. Define $\{T_n\}$ by $T_n = \frac{1}{\sqrt{n}}(Y_1+\ldots+Y_n)$, $n=1,2,\ldots$. Then $\frac{T_n}{\sqrt{n}} = \frac{1}{n}(Y_1+\ldots+Y_n) \to \frac{1}{\lambda}$ with probability one $[\lambda]$, so condition i) of Theorem 1.1 is met with $b(\lambda) = \frac{1}{\lambda}$. Also,

$$
\begin{aligned}
P_{\lambda_0}\{T_n > \sqrt{n}\, t\} &= P_{\lambda_0}\{(Y_1+\ldots+Y_n) > nt\} \\
&= \int_{nt}^{\infty} \frac{\lambda_0^n}{(n-1)!} x^{n-1} e^{-\lambda_0 x}\, dx \\
&= \int_{n\lambda_0 t}^{\infty} \frac{1}{(n-1)!} x^{n-1} e^{-x}\, dx \\
&= \sum_{k=0}^{n-1} \frac{(n\lambda_0 t)^k e^{-n\lambda_0 t}}{k!} \\
&= e^{-n\lambda_0 t}\left[1+n\lambda_0 t + \frac{(n\lambda_0 t)^2}{2!} + \ldots + \frac{(n\lambda_0 t)^{n-1}}{(n-1)!}\right] ,
\end{aligned}
$$

as can be seen from properties of gamma distributions (9) . Values of concern for t are those in a neighborhood of $\frac{1}{\lambda}$, where $\lambda < \lambda_0$, so it may be assumed that $\lambda_0 t > 1$. Thus each term in the sum above is larger than the preceding term, so

$$
e^{-n\lambda_0 t} \cdot \frac{(n\lambda_0 t)^{n-1}}{(n-1)!} < P_{\lambda_0}\{T_n > \sqrt{n}\, t\} < e^{-n\lambda_0 t} \cdot n \cdot \frac{(n\lambda_0 t)^{n-1}}{(n-1)!} .
$$

Using Stirling's formula, $\lim_{n\to\infty} \frac{\sqrt{2\pi(n-1)}\,(n-1)^{n-1}\, e^{-(n-1)}}{(n-1)!} = 1$, so

$$\lim_{n\to\infty} \frac{1}{n} \log[e^{-n\lambda_0 t} \frac{(n\lambda_0 t)^{n-1}}{(n-1)!}] = \lim_{n\to\infty} \frac{1}{n} \log[e^{-n\lambda_0 t} \frac{(n\lambda_0 t)^{n-1}}{\sqrt{2\pi(n-1)}\,(n-1)^{n-1} e^{-(n-1)}}]$$

$$= -\lambda_0 t + \lim_{n\to\infty} \frac{1}{n} \log[\frac{1}{\sqrt{2\pi(n-1)}} \left(\frac{n}{n-1}\right)^{n-1} (\lambda_0 te)^{n-1}]$$

$$= -\lambda_0 t - \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} \log 2\pi(n-1) + \lim_{n\to\infty} \frac{n-1}{n} \log \frac{n}{n-1} + \lim_{n\to\infty} \frac{n-1}{n} \log \lambda_0 te$$

$$= -\lambda_0 t + 0 + 0 + \log \lambda_0 te$$

$$= -\lambda_0 t + \log \lambda_0 t + 1 .$$

Similarly, since $\lim_{n\to\infty} \frac{1}{n} \log n = 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log[e^{-n\lambda_0 t} \cdot n \cdot \frac{(n\lambda_0 t)^{n-1}}{(n-1)!} = -\lambda_0 t + \log \lambda_0 t + 1 .$$

Hence ii) of Theorem 1.1 is satisfied with $f(t)=\lambda_0 t-\log \lambda_0 t - 1$,

so $c(\lambda) = 2f(b(\lambda)) = 2\left[\frac{\lambda_0}{\lambda} - \log \frac{\lambda_0}{\lambda} - 1\right]$.

For the purpose of comparison, another example is now given, which is essentially the same as one by Bahadur (1) .

<u>Example 1.2</u>: Let Y_i be distributed as in Example 1.1, and define $X_i = 1$ if $Y_i \geq 1$ and $X_i = 0$ if $Y_i < 1$. Then $P_\lambda\{X_i=1\} = P_\lambda\{Y_i\geq 1\} = e^{-\lambda}$, so H and A in Example 1.1 are equivalent to H: $P_\lambda\{X_i=1\} = p_0 = e^{-\lambda_0}$ and A: $P_\lambda\{X_i=1\} = p_\lambda = e^{-\lambda} > p_0$. With

$\{T_n^*\}$ defined by $T_n^* = \frac{1}{\sqrt{n}}(X_1+\ldots+X_n)$, it follows that $\frac{T_n^*}{\sqrt{n}} \to e^{-\lambda}$ with probability one $[\lambda]$, so $b^*(\lambda) = e^{-\lambda}$. From a result by Bahadur (4),

$$\frac{1}{n}\log P_{\lambda_0}\{T_n^* > \sqrt{n}\, t\} \to -t\log\left(\frac{t}{e^{-\lambda_0}}\right) - (1-t)\log\left(\frac{1-t}{1-e^{-\lambda_0}}\right) .$$

Thus $f^*(t) = t\log\left(\frac{t}{e^{-\lambda_0}}\right) + (1-t)\log\left(\frac{1-t}{1-e^{-\lambda_0}}\right)$,

so by Theorem 1.1

$$c^*(\lambda) = 2f^*(b^*(\lambda)) = 2\left[e^{-\lambda}\log\left(\frac{e^{-\lambda}}{e^{-\lambda_0}}\right) + (1-e^{-\lambda})\log\left(\frac{1-e^{-\lambda}}{1-e^{-\lambda_0}}\right)\right] .$$

By combining the results of Examples 1.1 and 1.2, it is seen that the relative efficiency of $\{T_n^*\}$ with respect to $\{T_n\}$ is

$$\phi(\lambda) = \frac{e^{-\lambda}\log\left(\frac{e^{-\lambda}}{e^{-\lambda_0}}\right) + (1-e^{-\lambda})\log\left(\frac{1-e^{-\lambda}}{1-e^{-\lambda_0}}\right)}{\frac{\lambda_0}{\lambda} - \log\frac{\lambda_0}{\lambda} - 1}$$

Repeated use of L'Hospital's Rule gives $\psi(\lambda_0) = \lim_{\lambda\to\lambda_0} \phi(\lambda)$ as

$$\psi(\lambda_0) = \frac{\lambda_0^2\, e^{-\lambda_0}}{1-e^{-\lambda_0}} .$$

These examples have immediate application to life-testing if Y_i is regarded as the life time of the ith article on test.

An Upper Bound for Exact Slopes

One of the most interesting and important aspects of Bahadur efficiency is the relationship between exact slopes and the Kullback-Liebler information function. The Kullback-Liebler information function will now be defined. Let $\{P_\theta\}$ be a family of probability measures on the real numbers indexed by a set Θ . For simplicity, suppose each P_θ is absolutely continuous with respect to Lebesque measure. That is, there are density functions f_θ such that $P_\theta(S) = \int_S f_\theta(x)dx$ for each $\theta\varepsilon\Theta$ and each measurable set S of real numbers. Then for θ_0 and θ belonging to Θ , the Kullback-Liebler information number is

$$K(\theta,\ \theta_0) = \underset{\theta}{E}[\log \frac{f_\theta(x)}{f_{\theta_0}(x)}].$$

Bahadur (3) shows that $K(\theta, \theta_0)$ provides an upper bound for exact slopes, and that the likelihood ratio statistic attains the upper bound. The result is formally presented without proof in the next theorem.

<u>Theorem 1.2</u>: If $\{T_n\}$ is a sequence of tests for the hypothesis $H\colon \theta\varepsilon\Theta_0$ versus $A\colon \theta\varepsilon\Theta-\Theta_0$ with exact slope $c(\theta)$, then for $\theta\varepsilon\Theta-\Theta_0$,

$$c(\theta) \leq 2J(\theta) = 2\inf_{\theta_0\varepsilon\Theta_0} K(\theta,\ \theta_0) .$$

If T_n is the likelihood ratio statistic, then equality holds.

An application of Theorem 1.2 is to determine if a test is close to optimal in the sense of asymptotic efficiency, as is illustrated in Example 5 of (14).

CHAPTER II

PARTIAL CHARACTERIZATIONS OF BAHADUR EFFICIENCY

Comparisons have been made between various criteria for judging between tests, and in particular, Bahadur efficiency has been compared to other measures of asymptotic efficiency (see (1) and (3)). Also, Bahadur (1) has shown that relations exist between Bahadur efficiency and a criterion called domination, which is now discussed.

Domination

For two-standard sequences $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ and for fixed θ and α , $0 < \alpha < 1$, let $\beta_n^{(i)}(\alpha|\theta) = P_\theta\{L^{(i)}(T_n^{(i)}) > \alpha\}$, $i=1,2$, and let $\delta_n(1,2|\theta) = \sup_\alpha[\beta_n^{(2)}(\alpha|\theta) - \beta_n^{(1)}(\alpha|\theta)]$. (Notice that $1-\beta_n^{(i)}(\alpha|\theta)$ is the power of a size α test which rejects for large values of $T_n^{(i)}$.) Then $\{T_n^{(2)}\}$ is said to <u>dominate</u> $\{T_n^{(1)}\}$ at θ if $\lim_{n\to\infty} \delta_n(1,2|\theta) = 0$.

Two of the relations between domination and Bahadur efficiency are given in two theorems that follow. For complete proofs, see (1).

<u>Theorem 2.1</u>: If $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$ at θ , then $\phi_{12}^{(a)}(\theta) \leq 1$.

<u>Theorem 2.2</u>: If $\phi_{12}^{(a)}(\theta) < 1$, then $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$ at θ.

It is asserted in (1) and in (7) that from Theorems 2.1 and 2.2 one can conclude that $\phi_{12}^{(a)}(\theta) < 1$ if and only if if $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$ but $\{T_n^{(1)}\}$ does not dominate $\{T_n^{(2)}\}$. However, this is not quite true. The "only if" portion of the assertion readily follows from the theorems, but the "if" portion does not. More specifically, assuming $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$ but $\{T_n^{(1)}\}$ does not dominate $\{T_n^{(2)}\}$, the most one can conclude from Theorems 2.1 and 2.2 is that $\phi_{12}^{(a)}(\theta) \leq 1$. To illustrate that this is the case, an example is now given of two standard sequences $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ for which $\phi_{12}^{(a)}(\theta) = 1$, $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$, but $\{T_n^{(1)}\}$ does not dominate $\{T_n^{(2)}\}$. Thus it will be seen that Bahadur efficiency does not distinguish between tests as sharply as previously thought, in that there are pairs of tests which are ordered by the domination criterion but not by Bahadur efficiency. It should be noted at this point that the example does not counter any statement which is proven in (1) or (7), but instead a statement of claim about consequences of true theorems.

<u>Example 2.1</u>: Let $X_1, X_2, \ldots,$ be distributed normally and independently with variance one and mean μ, where μ is equal either to zero or to a fixed and known value $m > 0$, and consider the hypotheses $H: \mu=0$ and $A: \mu=m$. For $n=1,2,\ldots$ take $T_n^{(2)} = \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$. Then $T_n^{(2)}$ is distributed normally with mean $\sqrt{n}\mu$ and variance unity for each n, so when $\mu = 0$, $F_n^{(2)}(t) = \Phi(t)$, where Φ is the distribution function of a normally distributed random variable with mean zero and unit variance. Thus $F_n(t) \to \Phi(t)$, so condition i) in the definition of a standard sequence is met. Also, Φ has the form prescribed by condition ii) of the definition with $a=1$, as is shown in (1). Finally,

$\frac{T_n^{(2)}}{\sqrt{n}} \to \mu$ in probability $[\mu]$, so condition iii) is met, and thus $\{T_n^{(2)}\}$ is a standard sequence with slope $c_2^{(a)}(m) = m^2$.

For fixed $\beta^{(2)}$, $0 < \beta^{(2)} < 1$, define the sequence $\{\alpha_n\}$ by $\beta_n^{(2)}(\alpha_n|m) = \beta^{(2)}$. Now take $\beta^{(1)}$ such that $\beta^{(2)} < \beta^{(1)} < 1$, and for all $n=1,2,\ldots$ for which $\beta_1^{(2)}(\alpha_n|m) \geq \beta^{(1)}$ (that is, for all n larger than some n_0), define k_n by

$$\beta_{k_n}^{(2)}(\alpha_n|m) \geq \beta^{(1)} > \beta_{k_n+1}^{(2)}(\alpha_n|m) .$$

Then define $\{T_n^{(1)}\}$ by $T_n^{(1)} = T_{k_n}^{(2)}$. Thus $F_n^{(1)} = F_{k_n}^{(2)}$, so it is immediately seen that $\{T_n^{(1)}\}$ satisfies i) and ii) of the definition of a standard sequence with $a=1$.

Now

$$\beta_n^{(2)}(\alpha|\mu) = \int_{-\infty}^{z_\alpha} N(x;\ \sqrt{n}\mu,1)dx$$

$$= \int_{-z_\alpha+\sqrt{n}\mu}^{\infty} N(x;0,1)dx ,$$

where $N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ and $\int_{z_\alpha}^{\infty} N(x;0,1)dx = \alpha$,

so

$$\beta_n^{(1)}(\alpha|m) = \beta_{k_n}^{(2)}(\alpha|m)$$

$$= \int_{-z_\alpha+\sqrt{k_n}m}^{\infty} N(x;0,1)dx .$$

Thus with L_i defined by $\beta^{(i)} = \int_{L_i}^{\infty} N(x;0,1)dx$, i=1,2, it follows that $L_2 = -z_{\alpha_n} + \sqrt{n}\, m$ and from the definition of $\{k_n\}$ it follows that

$$-z_{\alpha_n} + \sqrt{k_n}\, m \leq L_1 < -z_{\alpha_n} + \sqrt{k_n+1}\, m$$

for $n > n_0$. Now $\sqrt{k+1} - \sqrt{k} = \frac{1}{\sqrt{k+1} + \sqrt{k}} \to 0$ as $k \to \infty$, so $-z_{\alpha_n} + \sqrt{k_n}\, m \to L_1$, and hence

$$(\sqrt{k_n} - \sqrt{n})m = (-z_{\alpha_n} + \sqrt{k_n}\, m) - (-z_{\alpha_n} + \sqrt{n}\, m)$$

$$\to L_1 - L_2 \ .$$

Thus $\sqrt{k_n} - \sqrt{n} \to \frac{L_1 - L_2}{m}$, so $\sqrt{\frac{k_n}{n}} - 1 \to 0$, and therefore $\frac{k_n}{n} \to 1$. So it is clear that

$$\frac{T_n^{(1)}}{\sqrt{n}} = \frac{T_{k_n}^{(2)} / \sqrt{k_n}}{\sqrt{\frac{n}{k_n}}}$$

$$\to m$$

in probability [m] , and hence $\{T_n^{(1)}\}$ is a standard sequence with slope $c_1^{(a)}(m) = m^2$. Thus $\phi_{12}^{(a)}(m) = 1$.

Now $\beta_n^{(2)}(\alpha_n|m) = \beta^{(2)} < \beta^{(1)} \leq \beta_{k_n}^{(2)}(\alpha_n|m) = \beta_n^{(1)}(\alpha_n|m)$ for all $n > n_0$, so

$$\delta_n(2,1|m) = \sup_\alpha[\beta_n^{(1)}(\alpha|m) - \beta_n^{(2)}(\alpha|m)]$$

$$\geq \beta_n^{(1)}(\alpha_n|m) - \beta_n^{(2)}(\alpha_n|m)$$

$$\geq \beta^{(1)} - \beta^{(2)} ,$$

and therefore $\delta_n(2,1|m)$ does not have a limit of zero as $n \to \infty$. Thus $\{T_n^{(1)}\}$ does not dominate $\{T_n^{(2)}\}$. But $\beta_n^{(2)}(\alpha|m) \leq \beta_n^{(1)}(\alpha|m)$ for all $n > n_0$, so $\delta_n(1,2|m) = 0$ for $n > n_0$, which implies $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$.

Ratio of Levels

It was seen in the previous section that if two tests are ordered according to Bahadur efficiency, then they are also ordered according to the domination criterion. The next theorem shows that ordering by Bahadur efficiency also implies ordering in the sense that the level attained by one test statistic becomes infinitely smaller than the level attained by the other test statistic. The theorem is true for both approximate and exact efficiency, although it is given here in terms of approximate efficiency.

Theorem 2.3: If $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two standard sequences for which $\phi_{12}^{(a)}(\theta) < 1$, then $\dfrac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} \to 0$ in probability $[\theta]$.

Proof: Now $\phi_{12}^{(a)}(\theta) < 1$ implies $c_1^{(a)}(\theta) < c_2^{(a)}(\theta)$, and from Chapter I,

$$-\frac{2}{n} \log L^{(i)}(T_n^{(i)}) \to c_i^{(a)}(\theta)$$

in probability $[\theta]$. Thus it is seen that

$$- \frac{2}{n} \log \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} = - \frac{2}{n} \log L^{(2)}(T_n^{(2)}) + \frac{2}{n} \log L^{(1)}(T_n^{(1)})$$

$$\to c_2^{(a)}(\theta) - c_1^{(a)}(\theta)$$

in probability $[\theta]$. So for $\varepsilon > 0$, $k = c_2^{(a)}(\theta) - c_1^{(a)}(\theta)$, and $n > - \frac{4 \log \varepsilon}{k}$, it follows that

$$P_\theta \left\{ \left| \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} \right| < \varepsilon \right\} = P_\theta \left\{ \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} < \varepsilon \right\}$$

$$> P_\theta \left\{ \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} < e^{-\frac{n}{2}(\frac{k}{2})} \right\}$$

$$= P_\theta \left\{ - \frac{2}{n} \log \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} > \frac{k}{2} \right\}$$

$$= P_\theta \left\{ - \frac{2}{n} \log \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} - k > - \frac{k}{2} \right\}$$

$$> P_\theta \left\{ \left| - \frac{2}{n} \log \frac{L^{(2)}(T_n^{(2)})}{L^{(1)}(T_n^{(1)})} - k \right| < \frac{k}{2} \right\}$$

$$\to 1$$

and the theorem is established.

Notice that the contrapositive shows that if the ratio of levels converges to a constant (regardless of how large), then $\phi_{12}^{(a)} = 1$.

An example is now given to illustrate Theorem 2.3 and to emphasize even more strongly the fact that asymptotic results based on approximate levels may be misleading, as is discussed by Bahadur in (1). The statistics used in the example are those of Example 2 of (1).

Example 2.2: Let $X_1, X_2, \ldots$ be normally and independently distributed with mean zero and variance σ^2, and consider H: $\sigma^2 = 1$ and A: $\sigma^2 > 1$. As is shown in (1), with $T_n^{(2)} = \sqrt{2\Sigma X_i^2} - \sqrt{2n}$ and $T_n^{(3)} = \frac{1}{\sqrt{2n}} (\Sigma X_i^2 - n)$, the approximate Bahadur efficiency is given by

$$\phi_{23}^{(a)}(\sigma) = \frac{4}{[1+\sigma]^2} .$$

Thus $\phi_{23}^{(a)}(\sigma) < 1$, so by Theorem 2.3, $\frac{L^{(2)}(T_n^{(2)})}{L^{(3)}(T_n^{(3)})} \to \infty$ in probability $[\sigma]$. It should be noticed that for fixed n, both statistics are monotone functions of ΣX_i^2 and hence for practical purposes are equivalent, although both the relative efficiency and the limit of the ratio of levels indicate superiority of $T_n^{(3)}$ over $T_n^{(2)}$. This is a result of the fact that $L^{(2)} = L^{(3)}$, but $T_n^{(3)}$ is always larger than $T_n^{(2)}$. Closer analysis reveals this point, as follows.

Now

$$\Sigma X_i^2 = \frac{1}{2} (T_n^{(2)} + \sqrt{2n})^2 , \quad \text{so}$$

$$T_n^{(3)} = \frac{\frac{1}{2} (T_n^{(2)} + \sqrt{2n})^2 - n}{\sqrt{2n}}$$

$$= T_n^{(2)} + \frac{[T_n^{(2)}]^2}{2\sqrt{2n}} .$$

Also, $F^{(2)}(t) = F^{(3)}(t) = \Phi(t)$ (See (1)), so

$$\frac{L^{(2)}(T_n^{(2)})}{L^{(3)}(T_n^{(3)})} = \frac{1-\Phi(T_n^{(2)})}{1-\Phi(T_n^{(3)})}$$

$$= \frac{\int_{T_n^{(2)}}^{\infty} N(t;0,1)dt}{\int_{T_n^{(3)}}^{\infty} N(t;0,1)dt}$$

$$= \frac{\int_{t_n}^{\infty} N(t;0,1)dt}{\int_{[t_n+\frac{t_n^2}{2\sqrt{2n}}]}^{\infty} N(t;0,1)dt}$$

where, for the sake of notation, $t_n = T_n^{(2)}$. Now from (5) , for $x > 0$,

$$\frac{x^2-1}{x^3} N(x;0,1) < \int_x^{\infty} N(t;0,1)dt < \frac{1}{x} N(x;0,1) .$$

Hence

$$\frac{\int_{t_n}^{\infty} N(t;0,1)dt}{\int_{[t_n+\frac{t_n^2}{2\sqrt{2n}}]}^{\infty} N(t;0,1)dt} > \frac{\frac{t_n^2-1}{t_n^3}\, e^{-\frac{1}{2}t_n^2}}{\frac{1}{[t_n+\frac{t_n^2}{2\sqrt{2n}}]}\, e^{-\frac{1}{2}(t_n+\frac{t_n^2}{2\sqrt{2n}})^2}}$$

$$= t_n\left(1 + \frac{t_n}{2\sqrt{2n}}\right)\frac{t_n^2 - 1}{t_n^3}\, e^{-\frac{1}{2}\left[t_n^2 - t_n^2 - 2\cdot\frac{t_n^3}{2\sqrt{2n}} - \frac{t_n^4}{8n}\right]}$$

$$= \left(1 + \frac{t_n}{2\sqrt{2n}}\right)\frac{t_n^2 - 1}{t_n^2}\, e^{\left(\frac{t_n^3}{2\sqrt{2n}} + \frac{t_n^4}{16n}\right)} .$$

Now

$$\frac{t_n}{2\sqrt{2n}} = \frac{\sqrt{2\Sigma X_i^2} - \sqrt{2n}}{2\sqrt{2n}}$$

$$= \frac{1}{2}\left[\sqrt{\frac{\Sigma X_i^2}{n}} - 1\right]$$

$$\to \frac{1}{2}(\sigma - 1) \text{ in probability } [\sigma] .$$

Also, $\frac{t_n^2 - 1}{t_n^2} \to 1$ and $\left(\frac{t_n^3}{2\sqrt{2n}} + \frac{t_n^4}{16n}\right) \to \infty$ in probability $[\sigma]$,

so it follows that $\frac{L^{(2)}(T_n^{(2)})}{L^{(3)}(T_n^{(3)})} \to \infty$ in probability $[\sigma]$.

The preceding theorem reveals the asymptotic behavior of the ratio of levels when the efficiency is not unity. A theorem is now presented which provides a tool for examining the behavior of the ratio of levels when the efficiency is equal to one. First, however, notice that the equation

$$-\frac{2}{n}\log L(T_n) = c^{(a)}(\theta) + \varepsilon_n , \text{ where } \varepsilon_n \to 0 \text{ in probability } [\theta]$$

follows from the fact that the left hand side converges to $c^{(a)}(\theta)$ in probability $[\theta]$. (see (1)).

Theorem 2.4: For two standard sequences $\{T_n^{(i)}\}$ with equal slopes, write $-\frac{2}{n}\log L^{(i)}(T_n^{(i)}) = c^{(a)}(\theta) + \varepsilon_n^{(i)}$, $\varepsilon_n^{(i)} \to 0$ in probability $[\theta]$. Then $\frac{L^{(1)}(T_n^{(1)})}{L^{(2)}(T_n^{(2)})}$ converges to zero, to one, to infinity, or does not converge depending on whether $n(\varepsilon_n^{(1)} - \varepsilon_n^{(2)})$ converges to infinity, to zero, to negative infinity, or does not converge, respectively.

Proof: From $\frac{-2 \log L^{(i)}(T_n^{(i)})}{n} = c^{(a)}(\theta) + \varepsilon_n^{(i)}$, it is seen that

$$L^{(i)}(T_n^{(i)}) = e^{-\frac{n}{2}(c^{(a)}(\theta) + \varepsilon_n^{(i)})},$$

so

$$\frac{L^{(1)}(T_n^{(1)})}{L^{(2)}(T_n^{(2)})} = \frac{e^{-\frac{n}{2}(c^{(a)}(\theta) + \varepsilon_n^{(1)})}}{e^{-\frac{n}{2}(c^{(a)}(\theta) + \varepsilon_n^{(2)})}} = e^{-\frac{n}{2}(\varepsilon_n^{(1)} - \varepsilon_n^{(2)})}.$$

The conclusion of the theorem follows immediately.

Suppose $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two standard sequences which from all appearances are equally good. One may suspect that the ratio of their levels would converge to unity. However, Theorem 2.4 can be used to show this is not necessarily the case, as is illustrated in the next example.

Example 2.3: Let X_i be distributed normally and independently with mean μ and variance unity, and suppose the sample size $n = 2k$, $k = 1,2,\ldots$ is always even. Consider the hypotheses $H: \mu=0$ and $A: \mu > 0$. Let $T_n^{(1)} = \sqrt{k}\,(X_1+\ldots+X_k)$ and $T_n^{(2)} = \sqrt{k}\,(X_{k+1}+\ldots+X_n)$. Then $T_n^{(i)}$ is a standard sequence with slope $\frac{\mu^2}{2}$ whose distribution function is Φ under H, $i = 1, 2$.

Now, as in the proof of Theorem 2.4, $L(T_n) = e^{-\frac{n}{2}(c^{(a)}(\theta)+\varepsilon_n)}$ and from the definition of a standard sequence, $L(T_n) = e^{-\frac{1}{2}T_n^2[1+\delta(T_n)]}$, where $\lim_{t\to\infty} \delta(t) = 0$. Thus $\varepsilon_n = (\frac{T_n^2}{n} - c^{(a)}(\theta)) + \frac{T_n^2}{n}\,\delta(T_n)$. Hence

$$n(\varepsilon_n^{(1)} - \varepsilon_n^{(2)}) = [T_n^{2(1)} - T_n^{2(2)}] + [T_n^{2(1)}\,\delta(T_n^{(1)}) - T_n^{2(2)}\,\delta(T_n^{(2)})]\ ,$$

since $c_1^{(a)}(\mu) = c_2^{(a)}(\mu) = \frac{\mu^2}{2}$, and where $1 - \Phi(x) = e^{-\frac{1}{2}x^2[1+\delta(x)]}$.

By Theorem 2.4, if $\frac{L^{(1)}(T_n^{(1)})}{L^{(2)}(T_n^{(2)})} \to 1$, then $n(\varepsilon_n^{(1)} - \varepsilon_n^{(2)}) \to 0$.

From (5),

$$[\frac{x^2-1}{x^3}]\,\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}x^2} < e^{-\frac{1}{2}x^2[1+\delta(x)]} < \frac{1}{x}\cdot\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}x^2}\ .$$

So after some manipulation,

$$\log(\sqrt{2\pi}\ x) < \frac{1}{2}x^2\ \delta(x) < \log(\sqrt{2\pi}x^3/[x^2-1])\ .$$

Thus, for two real numbers $x > 0$ and $y > 0$,

$$\log\ \frac{x^3/[x^2-1]}{y} = \log\ \frac{\sqrt{2\pi}\ x^3/[x^2-1]}{\sqrt{2\pi}\ y}$$

$$= \log(\sqrt{2\pi}\; x^3/[x^2-1]) - \log(\sqrt{2\pi} y)$$

$$> \frac{1}{2} x^2\delta(x) - \frac{1}{2} y^2\delta(y)$$

$$> \log(\sqrt{2\pi}\; x) - \log(\sqrt{2\pi}\; y^3/[y^2-1])$$

$$= \log \frac{\sqrt{2\pi}\; x}{\sqrt{2\pi} y^3/[y^2-1]}$$

$$= - \log \frac{y^3/[y^2-1]}{x}$$

Therefore,

$$\log \frac{T_n^{3\,(1)}/[T_n^{2\,(1)}-1]}{T_n^{(2)}} > \frac{1}{2}\,[T_n^{2\,(1)}\delta(T_n^{(1)}) - T_n^{2\,(2)}\delta(T_n^{(2)})]$$

$$> - \log \frac{T_n^{3\,(2)}/[T_n^{2\,(2)} - 1]}{T_n^{(1)}} \; .$$

But

$$\frac{T_n^{3\,(1)}/[T_n^{2\,(1)} - 1]}{T_n^{(2)}} = \frac{\left(\frac{T_n^{(1)}}{\sqrt{n}}\right)^3 \Big/ \left[\left(\frac{T_n^{(1)}}{\sqrt{n}}\right)^2 - \frac{1}{n}\right]}{\frac{T_n^{(2)}}{\sqrt{n}}}$$

$$\rightarrow \frac{\left(\frac{\mu}{\sqrt{2}}\right)^3 \Big/ \left[\left(\frac{\mu}{\sqrt{2}}\right)^2 - 0\right]}{\frac{\mu}{\sqrt{2}}}$$

$$= 1 \; ,$$

so

$$\log \frac{T_n^{3\,(1)}/[T_n^{2\,(1)} - 1]}{T_n^{(2)}} \rightarrow 0$$

and similarly,

$$\log \frac{T_n^{3(2)}/[T_n^{2(2)} - 1]}{T_n^{(1)}} \to 0 \; .$$

Hence

$$T_n^{2(1)}\delta(T_n^{(1)}) - T_n^{2(2)}\delta(T_n^{(2)}) \to 0 \; .$$

Since

$$n(\varepsilon_n^{(1)} - \varepsilon_n^{(2)}) = T_n^{2(1)}\delta(T_n^{(1)}) - T_n^{2(2)}\delta(T_n^{(2)}) + T_n^{2(1)} - T_n^{2(2)} \; ,$$

it follows that $n(\varepsilon_n^{(1)} - \varepsilon_n^{(2)}) \to 0$ only if $T_n^{2(1)} - T_n^{2(2)} \to 0$, which in turn occurs only if $T_n^{(1)} - T_n^{(2)} \to 0$, because

$$T_n^{2(1)} - T_n^{2(2)} = (T_n^{(1)} + T_n^{(2)})(T_n^{(1)} - T_n^{(2)}) \; .$$

It will now be shown that $T_n^{(1)} - T^{(2)}$ does not converge to zero in probability, and hence that $\dfrac{L^{(1)}(T_n^{(1)})}{L^{(2)}(T_n^{(2)})}$ does not converge to unity in probability. To do so, let $\varepsilon < \dfrac{\sqrt{2\pi}}{2}$. Then

$$\begin{aligned}
P_\mu\{|T_n^{(1)} - T_n^{(2)}| < \varepsilon\} &= P_\mu\{T_n^{(2)} - \varepsilon < T_n^{(1)} < T_n^{(2)} + \varepsilon\} \\
&= E_\mu[P_\mu\{T_n^{(2)} - \varepsilon < T_n^{(1)} < T_n^{(2)} + \varepsilon \mid T_n^{(2)}\}] \\
&< E_\mu[2\varepsilon \cdot \frac{1}{\sqrt{2\pi}}] \\
&= \frac{2\varepsilon}{\sqrt{2\pi}} \\
&< 1 \; .
\end{aligned}$$

The first inequality holds because $T_n^{(1)}$ is normally distributed with variance one and is independent of $T_n^{(2)}$, and the fact that

$$\int_{K-\varepsilon}^{K+\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{n}(x-\nu)^2} dx < \int_{K+\varepsilon}^{K+\varepsilon} \frac{1}{\sqrt{2\pi}} dx$$

$$= \frac{2\varepsilon}{\sqrt{2\pi}}$$

for any K and ν. Thus for $\varepsilon < \frac{\sqrt{2\pi}}{2}$, $P_\mu\{|T_n^{(1)} - T_n^{(2)}| < \varepsilon\}$ is bounded by a constant smaller than one, and hence the probability cannot converge to one, and the desired conclusion is established.

One final example is now presented to show that two tests can be judged equivalent according to approximate Bahadur efficiency, and yet have the ratio of levels converge to infinity.

<u>Example 2.4</u>: Let X_i be distributed normally and independently with mean μ and variance one, and consider $H: \mu=0$ versus $A: \mu\neq0$. Define $T_n^{(1)}$ and $T_n^{(2)}$ by $T_n^{2(1)} = n\overline{X}^2$ and $T_n^{2(2)} = (n + \sqrt{n})\overline{X}^2$. Then $\frac{T_n^{(i)}}{\sqrt{n}} \to |\mu|$ in probability $[\mu]$, $i=1,2,$. Also, $n\overline{X}^2$ has a chi-square distribution with one degree of freedom when $\mu=0$, and $\frac{n+\sqrt{n}}{n} \to 1$, so both $T_n^{(1)}$ and $T_n^{(2)}$ converge in distribution to the square root of a chi-square with one degree of freedom. Hence from [ ], $c_1^{(a)}(\mu) = c_2^{(a)}(\mu) = \mu^2$, so $\phi_{12}^{(a)}(\mu) = 1$.

Now $F(t) = \int_{-t}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$, so $1-F(t) = 2[1-\Phi(t)]$.

Thus

$$2\,\frac{t^2-1}{t^3}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}t^2} < 1-F(t) < 2 \cdot \frac{1}{t}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}t^2}$$

so

$$\frac{L^{(1)}(T_n^{(1)})}{L^{(2)}(T_n^{(2)})} = \frac{1-F(T_n^{(1)})}{1-F(T_n^{(2)})}$$

$$> \frac{\left(\frac{T_n^{2(1)} - 1}{T_n^{3(1)}}\right) e^{-\frac{1}{2} T_n^{2(1)}}}{\frac{1}{T_n^{(2)}} e^{-\frac{1}{2} T_n^{2(2)}}}$$

$$= \frac{\frac{n\overline{X}^2 - 1}{n^{3/2}\overline{X}^3}}{\frac{1}{\sqrt{n+\sqrt{n}}\ \overline{X}}} e^{-\frac{1}{2}[n\overline{X}^2 - (n\overline{X}^2 + \sqrt{n}\ \overline{X}^2]}$$

$$= \sqrt{\frac{n+\sqrt{n}}{n}} \cdot \frac{\overline{X}^2 - \frac{1}{n}}{\overline{X}^2} e^{\frac{1}{2}\sqrt{n}\overline{X}^2}$$

$$\to \infty$$

because $\sqrt{\frac{n+\sqrt{n}}{n}} \to 1$, $\frac{\overline{X}^2 - \frac{1}{n}}{\overline{X}^2} \to 1$ in probability one [μ] , and $\sqrt{n}\ \overline{X}^2 \to \infty$ in probability [μ] .

CHAPTER III

SOME INFORMATION THEORETIC PROPERTIES OF EXACT SLOPES

As was stated in Theorem 1.2, there is an intimate relation between exact slopes and the Kullback-Liebler information function. If $\{T_n\}$ is the sequence of likelihood ratio statistics for testing the hypothesis H: $\theta\varepsilon\Theta_0$ versus the alternative A: $\theta\varepsilon\Theta-\Theta$, then the exact slope of $\{T_n\}$ is given by

$$c(\theta) = 2 \inf_{\theta_0 \varepsilon \Theta_0} K(\theta, \theta_0) ,$$

where $K(\theta, \theta_0)$ is the Kullback-Liebler information function. This fact suggests that exact slopes in general may have some of the properties of information functions.

Some Desirable Properties Of Information Functions

Assume X is a random variable distributed according to a probability distribution P_θ which is indexed by a parameter $\theta\varepsilon\Theta$. There are various functional measures of the amount of information about θ in an observation X ; that is, functions which are calculated according to some formula involving θ and the family of distributions $\{P_\theta\}$. Such a function hopefully reflects the quality of inference about θ which can be made from X. Following are some generally accepted desirable properties for an information function, which will be denoted by $I_X(\theta)$ (see (12))

i) Nonnegativity: $I_X(\theta) \geq 0$ for all $\theta \varepsilon \Theta$.

ii) Invariance under parametric transformations: Let Θ and ϕ be two parameter spaces which both index the family of probability measures, and let g be a 1-1 transformation from Θ onto Φ . If $I_X(\theta)$ and $J_X(\phi)$ are two information functions calculated from the same algorithm with respect to Θ and Φ respectively, then $I_X(\theta) = J_X(g(\theta))$. (This property is possessed by the Kullback-Liebler information function but not by the Fisherian information function, which is defined by

$$F_X(\theta) = \underset{\theta}{E}[\frac{\partial}{\partial\theta} \ln f(x,\theta)]^2$$

iii) No information in a random variable whose distribution is independent of θ: If $P_\theta(X \leq x)$ is independent of θ , then $I_X(\theta) = 0$.

iv) No increase in information by data manipulation: If T is a function of X , then $I_T(\theta) \leq I_X(\theta)$.

v) Additivity of information in independent observations: If X and Y are two independent random variables, then $I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta)$.

Exact Slopes As An Asymptotic Measure Of Information

It was seen in Chapter I that if the exact slope $c_1(\theta)$ of a sequence of tests $\{T_n^{(1)}\}$ is smaller than the exact slope $c_2(\theta)$ of a sequence of tests $\{T_n^{(2)}\}$, then $\{T_n^{(2)}\}$ is judged superior to $\{T_n^{(1)}\}$ on the basis that the sequence of levels $\{L_n^{(2)}(T_n^{(2)})\}$ attained by $\{T_n^{(2)}\}$ converges to zero at a more rapid exponential rate than the sequence of levels $\{L_n^{(1)}(T_n^{(1)})\}$ attained by $\{T_n^{(1)}\}$. In this sense exact slopes reflect the quality of inference that may be made concerning the truth or falsity of the null hypothesis, and thus may be considered an asymptotic measure of the amount of information in the sequence of tests about the parameter θ. To further substantiate this claim, it will now be shown that exact slopes possess properties analogous to the desirable properties of information functions given in the previous section.

In developing these properties, restrict the tests to be continuous. Then for $\theta\varepsilon\Theta_0$ the levels are distributed uniformly on the interval (0,1).

i) Nonnegativity: The inequality $\dfrac{-2 \log L_n(T_n)}{n} > 0$ holds because $L_n(T_n) < 1$. Hence

$$c(\theta) = \lim_{n\to\infty} \frac{-2 \log L_n(T_n)}{n} \geq 0 .$$

ii) Invariance under parametric transformations: Denote by $\{P_\theta\}$, $\theta\varepsilon\Theta$, the family of probability measures indexed by $\theta\varepsilon\Theta$, and let $\{P_\theta\}$ be reparametrized by $\phi = g(\theta)$, where ϕ belongs to a new parameter space Φ and g is a one-to-one mapping from Θ onto Φ. That is, the family $\{P_\theta\}$, $\theta\varepsilon\Theta$, may be written $\{Q_\phi\}$, $\phi\varepsilon\Phi$, where $\phi = g(\theta)$ implies $Q_\phi(A) = P_\theta(A)$ for each measurable set A. Suppose $c(\theta)$ is the exact slope of a sequence of tests $\{T_n\}$ with respect to the parameter space Θ . This means that

$$P_\theta\{-\frac{2}{n}\log L_n \to c(\theta)\} = 1 .$$

From this equation, it may be concluded as follows that $\{T_n\}$ has an exact slope with respect to the parameter space Φ , and that it is invariant with respect to the parametric transformation g . Let $c^*(\phi)=c(g^{-1}(\phi))$. Then

$$\begin{aligned} Q_\phi\{-\frac{2}{n}\log L_n \to c^*(\phi)\} &= Q_\phi\{-\frac{2}{n}\log L_n \to c[g^{-1}(\phi)\} \\ &= Q_{g(\theta)}\{-\frac{2}{n}\log L_n \to c[g^{-1}(g(\theta))]\} \\ &= P_\theta\{-\frac{2}{n}\log L_n \to c(\theta)\} \\ &= 1 , \end{aligned}$$

with $\phi = g(\theta)$. Hence $c^*(\phi)$ is the exact slope of $\{T_n\}$ with respect to Φ , and $c(\theta) = c^*(g(\theta))$.

iii) No information in sequence of tests $\{T_n\}$ whose distribution does not depnd on θ: It will be shown that if the distribution of T_n is independent of θ , then $P_\theta\{-\frac{2}{n}\log L_n \to 0\} = 1$ for all θ , and hence $c(\theta) \equiv 0$. For simplicity, write $L_n(T_n(s)) = L_n(s)$.

Now if the sequence s is such that $-\frac{2}{n}\log L_n(s)$ does not converge to zero, then there is an $\varepsilon_s > 0$ such that $-\frac{2}{n}\log L_n(s) \geq \varepsilon_s$ for infinitely many n . For $\varepsilon > 0$ and positive integers m , define $A_{m,\varepsilon}$ by

$$A_{m,\varepsilon} = \{s: -\frac{2}{n}\log L_n(s) \geq \varepsilon \text{ for some } n \geq m\}$$

It follows that $s\varepsilon A_{m,\varepsilon_s}$ for each m. Hence $s\varepsilon A_{\varepsilon_s}$, where, for $\varepsilon > 0$, $A_\varepsilon = \bigcap_{m=1}^{\infty} A_{m,\varepsilon}$. Now $P_\theta(A_\varepsilon) = \lim_{m\to\infty} P_\theta(A_{m,\varepsilon})$ because $\{A_{m,\varepsilon}\}$ is a decreasing sequence in m. Also, $A_{m,\varepsilon} = \bigcup_{n=m}^{\infty} \{s: -\frac{2}{n}\log L_n(s) \geq \varepsilon\}$, so

$$P_\theta\{A_{m,\varepsilon}\} \leq \sum_{n=m}^{\infty} P_\theta\{s: -\frac{2}{n}\log L_n(s) \geq \varepsilon\}$$

$$= \sum_{n=m}^{\infty} P_\theta\{s: L_n(s) \leq e^{-\frac{n}{2}\varepsilon}\}$$

But L_n is distributed uniformly on $(0,1)$ for all θ, so

$$P_\theta\{A_{m,\varepsilon}\} \leq \sum_{n=m}^{\infty} e^{-\frac{n}{2}\varepsilon}$$

$$= \sum_{n=m}^{\infty} (e^{-\frac{\varepsilon}{2}})^n$$

$$= \frac{e^{-\frac{m\varepsilon}{2}}}{1-e^{-\frac{\varepsilon}{2}}} .$$

Thus $P_\theta\{A_\varepsilon\} = \lim_{m\to\infty} P_\theta\{A_{m,\varepsilon}\} = \lim_{m\to\infty} \frac{e^{-(m\varepsilon/2)}}{1-e^{-(\varepsilon/2)}} = 0$ for each $\varepsilon > 0$.

Now $\{s: -\frac{2}{n}\log L_n(s)$ does not converge to zero$\} \subset \bigcup_{\varepsilon>0} A_\varepsilon$, and $\bigcup_{\varepsilon>0} A_\varepsilon = \bigcup_{n=1}^{\infty} A_{\varepsilon_n}$, where $\varepsilon_n = \frac{1}{n}$, so

$$P_\theta\{s: -\frac{2}{n}\log L_n(s) \text{ does not converge to zero}\} \leq P_\theta\{\bigcup_{\varepsilon>0} A_\varepsilon\}$$

$$= P_\theta\{\bigcup_{k=1}^{\infty} A_{\varepsilon_k}\}$$

$$\leq \sum_{k=1}^{\infty} P_\theta\{A_{\varepsilon_k}\}$$

$$= 0 .$$

Hence

$$P_\theta\{s: -\frac{2}{n}\log L_n(s) \to 0\}$$
$$= 1-P_\theta\{s: -\frac{2}{n}\log L_n(s) \text{ does not converge to zero}\}$$
$$= 1 \quad \text{for all } \theta\varepsilon\Theta ,$$

and therefore $c(\theta) \equiv 0$.

iv) No increase in slopes by data manipulation: It will be shown that if $\{T_n\}$ is a sequence of tests with exact slope $c(\theta)$ and if $\{g_n\}$ is a sequence of functions which obey a certain restriction, then the sequence $\{T_n^*\}$, where $T_n^* = g_n(T_n)$, has exact slope $c^*(\theta) \leq c(\theta)$.

Assume there exists a sequence of positive real numbers $\{t_n\}$, $t_n \to \infty$ as $n \to \infty$, such that

$$P_\theta\{s: \text{there exists } N_s \text{ such that } T_n(s) > t_n \text{ for } n > N_s\} = 1$$

for all $\theta\varepsilon\Theta-\Theta_0$,

and assume that the function $g_n(x)$ is monotone increasing for $x > t_n$. With these assumptions, sequences s which yield a large value of $T_n(s)$ also yield a large value of $T_n^*(s)$. Thus T_n^* would tend to reject for the same values of s for which T_n rejects, and perhaps some other values. An example of this concept is given in the Appendix.

Now let s' be a fixed sequence such that $T_n(s') > t_n$ for $n > N_{s'}$. Then $n > N_{s'}$ implies

$$\begin{aligned} L_n^*(s') &= P_0\{s: T_n^*(s) \geq T_n^*(s')\} \\ &= P_0\{s: g_n(T_n(s)) \geq g_n(T_n(s'))\} \\ &= P_0\{s: g_n(T_n(s)) \geq g_n(T_n(s')) \text{ and } T_n(s) > t_n\} \\ &+ P_0\{s: g_n(T_n(s)) \geq g_n(T_n(s')) \text{ and } T_n(s) \leq t_n\} \end{aligned}$$

Since $n > N_{s'}$, it follows that $T_n(s') > t_n$. Hence the condition $T_n(s) > t_n$ and $g_n(T_n(s)) \geq g_n(T_n(s'))$ is equivalent to $T_n(s) \geq T_n(s')$ because g_n is monotone for values larger than t_n. Thus

$$\begin{aligned} L_n^*(s') &= P_0\{s: T_n(s) \geq T_n(s')\} \\ &+ P_0\{s: g_n(T_n(s)) \geq g_n(T_n(s')) \text{ and } T_n(s) \leq t_n\} \\ &\leq P_0\{s: T_n(s) \geq T_n(s')\} \\ &= L_n(s') . \end{aligned}$$

Hence

$$\begin{aligned} c^*(\theta) &= \lim_{n\to\infty} -\frac{2}{n} \log L_n^*(s') \\ &\leq \lim_{n\to\infty} -\frac{2}{n} \log L_n(s') \\ &= c(\theta) . \end{aligned}$$

v) Additivity of slopes of independent sequences: Let $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ be two sequences of test statistics for the hypotheses H: $\theta\varepsilon\Theta_0$ and A: $\theta\varepsilon\Theta-\Theta_0$ which have exact slopes $c_1(\theta)$ and $c_2(\theta)$, respectively. It is desired to measure the combined information in $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$. Recall that exact slopes are based upon the rates of convergence for a fixed s' of the probabilities

$$L_n(T_n(s')) = P_0\{s: T_n(s) > T_n(s')\}$$

$$= P_0\{\text{obtaining a value of } T_n \text{ larger than the observed } T_n(s')\}.$$

Thus it would appear that the combined information in $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ might be measured by the rate of convergence for a fixed s' of

$$P_0\{\text{obtaining values of } T_n^{(1)} \text{ and } T_n^{(2)} \text{ larger than the observed values } T_n^{(1)}(s') \text{ and } T_n^{(2)}(s')\}$$

$$= P_0\{s: T_n^{(1)}(s)>T_n^{(1)}(s') \text{ and } T_n^{(2)}(s) > T_n^{(2)}(s')\}$$

$$= P_0\{s: T_n^{(1)}(s)>T_n^{(1)}(s')\}P_0\{s:T_n^{(2)}(s)>T_n^{(2)}(s')\}$$

$$= L_n^{(1)}(T_n^{(1)}(s')) \cdot L_n^{(2)}(T_n^{(2)}(s'))$$

So if one defines

$$L_n^{(1,2)}(T_n^{(1)}(s), T_n^{(2)}(s)) = L_n^{(1)}(T_n^{(1)}(s)) \cdot L_n^{(2)}(T_n^{(2)}(s)) ,$$

it follows immediately that with probability one $[\theta]$,

$$-\frac{2}{n}\log L_n^{(1,2)}(T_n^{(1)}(s),T_n^{(2)}(s)) = -\frac{2}{n}\log L_n^{(1)}(T_n^{(1)}(s))L_n^{(2)}(T_n^{(2)}(s))$$

$$= -\frac{2}{n}\log L_2^{(1)}(T_n^{(1)}(s)) - \frac{2}{n}\log L_n^{(2)}(T_n^{(2)}(s))$$

$$\to c_1(\theta) + c_2(\theta) .$$

Hence in this sense exact slopes are additive for independent tests. However, to remain in the true framework of exact slopes, there must be one sequence of tests $\{T_n\}$ from which the exact slopes are calculated. Perhaps the most common procedure for combining independent tests, sometimes called Fisher's method, relies on the fact that if a random variable U is distributed uniformly on (0,1), then -2 log U is distributed as a chi-square with two degrees of freedom. Hence, for $\theta\varepsilon\Theta_0$, the statistic

$$-2\log L_n^{(1)}(T_n^{(1)})L_n^{(2)}(T_n^{(2)}) = -2\log L_n^{(1)}(T_n^{(1)}) -2\log L_n^{(2)}(T_n^{(2)})$$

is distributed as a chi-square with four degrees of freedom. It will now be shown that exact slopes are additive under this method of combining.

Define the sequence $\{T_n\}$ by

$$T_n = \frac{-2}{\sqrt{n}}\log L_n^{(1)}(T_n^{(1)}) \cdot L_n^{(2)}(T_n^{(2)}) .$$

Then

$$\frac{T_n}{\sqrt{n}} = -\frac{2}{n}\log L_n^{(1)}(T_n^{(1)}) \cdot L_n^{(2)}(T_n^{(2)})$$

$$= \frac{-2\log L_n^{(1)}(T_n^{(1)})}{n} + \frac{-2\log L_n^{(2)}(T_n^{(2)})}{n}$$

$$\to c_1(\theta) + c_2(\theta) \text{ with probability one } [\theta] .$$

Thus condition i) of Theorem 1.1 is met with $b(\theta)=c_1(\theta)+c_2(\theta)$. To obtain condition ii) of Theorem 1.1, a lemma is needed concerning the distribution of the product of two uniformly distributed random variables.

Lemma 3.1: If X and Y are two independent random variables, each distributed uniformly on the interval (0,1) then the distribution of their product $Z = X \cdot Y$ is given by

$$P\{Z < z\} = z(1-\log z) .$$

Proof: Now

$$P\{Z < z\} = P\{XY < z\}$$

$$= P\{X < \tfrac{z}{Y}\}$$

$$= \underset{Y}{E}\{P\{X < \tfrac{z}{Y}|Y\}$$

$$= \int_0^z P\{X < \tfrac{z}{y}|y\}dy + \int_z^1 P\{X < \tfrac{z}{y}|y\} \, dy ,$$

Now, for $y < z$, $P\{X < \frac{z}{y}\} = 1$, and for $z < y$, $P\{X < \frac{z}{y}\} = \frac{z}{y}$. Hence, recalling the independence of X and Y ,

$$P\{Z < z\} = \int_0^z 1 \cdot dy + \int_z^1 \frac{z}{y} \, dy$$

$$= z + z(\log y \Big|_{y=z}^{y=1})$$

$$= z - z \log z$$

$$= z(1-\log z) ,$$

and the lemma is proved.

Thus

$$\frac{1}{n} \log P_0\{T_n > \sqrt{n}t\} = \frac{1}{n} \log P_0\left\{\frac{-2 \log L_n^{(1)}(T_n^{(1)})L_n^{(2)}(T_n^{(2)})}{\sqrt{n}} > \sqrt{n}t\right\}$$

$$= \frac{1}{n} \log P_0\{L_n^{(1)}(T_n^{(1)})\ L_n(T_n^{(2)}) < e^{-\frac{n}{2}t}\}$$

$$= \frac{1}{n} \log[e^{-\frac{nt}{2}}(1 + \frac{nt}{2})]$$

$$= -\frac{t}{2} + \frac{1}{n} \log (1 + \frac{nt}{2})$$

$$\rightarrow -\frac{t}{2} \text{ as } n \to \infty .$$

Therefore condition ii) of Theorem 1.1 is met with $f(t) = \frac{t}{2}$. So, using Theorem 1.1, it follows that

$$c(\theta) = 2f(b(\theta))$$

$$= 2 \cdot \frac{c_1(\theta) + c_2(\theta)}{2}$$

$$= c_1(\theta) + c_2(\theta) ,$$

which establishes the desired result that exact slopes are additive when the tests are combined using the method of Fisher.

It will be seen in the next chapter that the additivity of exact slopes continues to hold when any number of independent tests are combined using this method.

CHAPTER IV

EXACT SLOPES OF COMBINED TESTS

In this chapter the concept developed in Chapter III of calculating exact slopes of combined independent tests will be extended to include tests of different hypotheses, different methods of combining, and tests based on unequal sample sizes.

Notation and Setting

Let $S^{(1)}, \ldots, S^{(p)}$ be p sample spaces from which samples $s^{(1)} = (x_1^{(1)}, \ldots, \text{ad inf}), \ldots, s^{(p)} = (x_1^{(p)}, \ldots, \text{ad inf})$ are observed, and let the probability measures $\{P_{\theta_i}^{(i)}\}$ $i=1,\ldots,p$, defined on $S^{(i)}$, $i=1,\ldots,p$ respectively, be indexed by parameters θ_i, $i=1,\ldots,p$, which belong to parameter spaces $\Theta^{(i)}$, $i=1,\ldots,p$. Also let $\{T_n^{(i)}\}$, $i=1,\ldots,p$, be p sequences of test statistics for the null hypotheses $H^{(i)}: \theta_i \varepsilon \Theta_0^{(i)}$, $i=1,\ldots,p$, where $\Theta_0^{(i)} \subset \Theta^{(i)}$, $i=1,\ldots,p$. Denote by $L_n^{(i)}(s^{(i)})$, $i=1,\ldots,p$, the levels attained by $T_n^{(i)}$ when $s^{(i)}$ is observed from $S^{(i)}$, and suppose $\{T_n^{(i)}\}$ has exact slope $c_i(\theta_i)$, $i=1,\ldots,p$. That is, suppose

$$- \frac{2}{n} \log L_n^{(i)}(s^{(i)}) \to c_i(\theta_i)$$

with probability one $[\theta_i]$, $i=1,\ldots,p$.

It will be the objective of the next section to test the combined null hypothesis, obtained by forming the cross-product of the individual null parameter spaces, utilizing a sequence of tests formed by combining the $\{T_n^{(i)}\}$, $i=1,\ldots,p$. In symbols, define $\Theta=\Theta^{(1)} x \ldots x\ \Theta^{(p)}$ and $\Theta_0 = \Theta_0^{(1)} x \ldots x\ \Theta_0^{(p)}$. Then the sequences $\{T_n^{(i)}\}$, $i=1,\ldots,p$, will be combined to form a test of $H: \theta=(\theta_1,\ldots,\theta_p)\varepsilon\Theta_0$ versus $A: \theta\varepsilon\Theta-\Theta_0$.

Fisher's Method of Combining Independent Tests

The sequence $\{T_n\}$ of combined tests using Fisher's method is obtained by defining

$$T_n = \sqrt{-2 \log L_n^{(1)} \cdot L_n^{(2)} \cdot \ldots L_n^{(p)}} \ .$$

The exact slope $c(\theta)$ will be calculated using Theorem 1.1. First,

$$\frac{T_n}{\sqrt{n}} = \sqrt{-\frac{2}{n} \log L_n^{(1)} \ldots L_n^{(p)}}$$

$$= \sqrt{-\frac{2}{n} \log L_n^{(1)} - \ldots - \frac{2}{n} \log L_n^{(p)}}$$

$$\rightarrow \sqrt{c_1(\theta_1) + \ldots + c_p(\theta_p)}$$

with probability one $[\theta]$. Thus $b(\theta) = \sqrt{c_1(\theta_1) + \ldots + c_p(\theta_p)}$. To calculate $f(t)$ for part ii) of Theorem 1.1, a result from (1) , is used concerning the form of the distribution function of a random variable distributed as the square root of a chi-square. For a proof, see (1) .

Lemma 4.1: If $F(x) = P(X < x)$, where X is distributed as the square root of a chi-square with k degrees of freedom, then F has the form

$$\log[1 - F(x)] = -\frac{x^2}{2}[1+o(1)] \quad \text{as } x \to \infty .$$

Now, for $\theta \varepsilon \Theta_0$, that is, for $\theta_i \varepsilon \Theta_0^{(i)}$, $i=1,\ldots,p$, T_n is distributed as the square root of a chi square random variable with $2p$ degrees of freedom. Hence, for $t > 0$,

$$
\begin{aligned}
-\frac{1}{n}\log P_0\{T_n > \sqrt{n}t\} &= -\frac{1}{n}\log[1-F(\sqrt{n}t)] \\
&= -\frac{1}{n}\left(-\frac{nt^2}{2}[1+o(1)]\right) \quad \text{as } n \to \infty \\
&= \frac{t^2}{2}[1+o(1)] \quad \text{as } n \to \infty \\
&\to \frac{t^2}{2} .
\end{aligned}
$$

Thus $f(t) = \frac{t^2}{2}$, and therefore, by Theorem 1.1,

$$
\begin{aligned}
c(\theta) &= 2f(b(\theta)) \\
&= 2 \cdot \frac{[\sqrt{c_1(\theta_1) + \ldots + c_p(\theta_p)}]^2}{2} \\
&= c_1(\theta_1) + \ldots + c_p(\theta_p) .
\end{aligned}
$$

That is, the slope of the test combined by Fisher's method is the sum of the slopes of the individual tests.

Combined Test Based On Maximum Level

Consider the test procedure of rejecting H_0 if all the levels are smaller than some specified quantity. This suggests that the significance level should be the maximum level. This procedure may be analytically expressed as the sequence of tests $\{T_n^*\}$, where

$$T_n^* = -\frac{2}{n}\log \max_i L_n^{(i)} .$$

Theorem 1.1 will again be used to calculate $c^*(\theta)$, the exact slope of $\{T_n^*\}$.

It follows easily that $b^*(\theta) = \min_i c_i(\theta_i)$, because

$$\frac{T_n^*}{\sqrt{n}} = -\frac{2}{n}\log \max_i L_n^{(i)}$$

$$= -\frac{2}{n}\max_i \log L_n^{(i)}$$

$$= \min_i \left(-\frac{2}{n}\log L_n^{(i)}\right)$$

$$\to \min_i c_i(\theta_i) \text{ with probability one } [\theta] .$$

Another lemma will be utilized to calculate $f^*(t)$ for ii) of Theorem 1.1.

<u>Lemma 4.2</u>: If X_i, $i=1,\ldots,p$, are distributed independently and exponentially with parameter λ, that is, $P_\lambda\{X_i > t\} = e^{-\lambda t}$, $i=1,\ldots,p$, then the smallest order statistic $Y_1 = \min_i X_i$ is

distributed exponentially with parameter $p\lambda$.

Proof: Since $P_\lambda\{X_i > t\} = e^{-\lambda t}$, $i=1,\ldots,p$, it follows that

$$\begin{aligned} P_\lambda\{Y_1 > t\} &= P_\lambda\{\min_i X_i > t\} \\ &= P_\lambda\{X_i > t,\ i=1,\ldots,p\} \\ &= \prod_{i=1}^{p} P_\lambda\{X_i > t\} \\ &= \prod_{i=1}^{p} e^{-\lambda t} \\ &= e^{-p\lambda t} \end{aligned}$$

which yields the desired result.

Now, under H_0, $-2 \log L_n^{(i)}$ is distributed as a chi-square with two degrees of freedom, which is exponential with parameter $\lambda = \frac{1}{2}$. Thus, under H_0,

$$\sqrt{n}\, T_n^* = \min_i(-2 \log L_n^{(i)})$$

is distributed as the smallest order statistic from a random sample of size p from an exponential distribution with parameter $\lambda = \frac{1}{2}$. So, by Lemma 4.2,

$$\begin{aligned} -\frac{1}{n} \log P_0\{T_n^* > \sqrt{n}t\} &= -\frac{1}{n} \log P_0\{\sqrt{n}\, T_n^* > nt\} \\ &= -\frac{1}{n} \log e^{\frac{-pnt}{2}} \\ &= \frac{pt}{2}. \end{aligned}$$

Therefore, $f^*(t) = \frac{pt}{2}$, so by Theorem 1.1,

$$c^*(\theta) = 2f^*(b^*(\theta))$$

$$= 2\,\frac{p \cdot \min_i c_i(\theta_i)}{2}$$

$$= p \cdot \min_i c_i(\theta_i) \;.$$

That is, the slope of the combined test based on the maximum level is the number of tests times the minimum of the individual slopes.

Combined Test Based on Minimum Level

As a third test procedure, consider using the minimum level. The sequence of tests $\{\tilde{T}_n\}$ analytically expresses the procedure, where

$$\tilde{T}_n = \frac{-2}{\sqrt{n}} \log \min_i L_n^{(i)} \;.$$

Once again, Theorem 1.1 is employed to calculate $\tilde{c}(\theta)$, the exact slope of $\{\tilde{T}_n\}$.

Now

$$\frac{\tilde{T}_n}{\sqrt{n}} = -\frac{2}{n} \log \min_i L_n^{(i)}$$

$$= -\frac{2}{n} \min_i \log L_n^{(i)}$$

$$= \max_i\left(-\frac{2}{n} \log L_n^{(i)}\right)$$

$$\to \max_i c_i(\theta_i) \text{ with probability one } [\theta] .$$

Thus i) of Theorem 1.1 is satisfied with $\tilde{b}(\theta) = \max_i c_i(\theta_i)$. To calculate $\tilde{f}(t)$ for ii) of Theorem 1.1, write

$$\begin{aligned}
P_0\{\tilde{T}_n > \sqrt{n}t\} &= P_0\{\sqrt{n}\,\tilde{T}_n > nt\} \\
&= P_0\{-2 \log \min_i L_n^{(i)} > nt\} \\
&= P_0\{\max_i(-2 \log L_n^{(i)}) > nt\} \\
&= P_0\{-2 \log L_n^{(i)} > nt \text{ for at least one } i\} \\
&= 1-P_0\{-2 \log L_n^{(i)} \leq nt \text{ for all } i,\ i=1,\ldots,p\} \\
&= 1-(1-e^{-nt/2})^p \\
&= 1 - \sum_{j=0}^{p} (-1)^j \binom{p}{j} e^{-\frac{jnt}{2}} \\
&= \binom{p}{1} e^{-nt/2} - \binom{p}{2} e^{-\frac{2nt}{2}} - \cdots - (-1)^p \binom{p}{p} e^{-pnt} .
\end{aligned}$$

Now for $j > 1$,

$$\frac{\binom{p}{j} e^{-\frac{jnt}{2}}}{\binom{p}{1} e^{-\frac{nt}{2}}} = \frac{\binom{p}{j}}{\binom{p}{1}} e^{-\frac{n(j-1)t}{2}}$$

$$\to 0 \quad \text{as} \quad n \to \infty$$

so

$$\frac{\sum_{j=2}^{p} (-1)^j \binom{p}{j} e^{-\frac{jnt}{2}}}{\binom{p}{1} e^{-\frac{nt}{2}}} \to 0 \quad \text{as} \quad n \to \infty .$$

Therefore,

$$-\sum_{j=2}^{p} (-1)^j \binom{p}{j} e^{-\frac{jnt}{2}} = \binom{p}{1} e^{-\frac{nt}{2}} o(1) \text{ as } n \to \infty .$$

Hence

$$P_0\{\tilde{T}_n > \sqrt{n}t\} = \binom{p}{1} e^{-\frac{nt}{2}} [1+o(1)] \quad \text{as} \quad n \to \infty ,$$

so

$$-\frac{1}{n} \log P_0\{\tilde{T}_n > \sqrt{n}t\} = -\frac{1}{n} \log \binom{p}{1} + \frac{t}{2}[1+o(1)] \text{ as } n \to \infty$$

$$\to \frac{t}{2} .$$

Thus $\tilde{f}(t) = \frac{t}{2}$, and Theorem 1.1 gives

$$\tilde{c}(\theta) = 2 \cdot \tilde{f}(\tilde{b}(\theta))$$

$$= 2 \cdot \frac{\max_i c_i(\theta_i)}{2}$$

$$= \max_i c_i(\theta_i) .$$

That is, the slope of the combined test based on the minimum level is equal to the maximum of the individual slopes.

Combining Tests With Unequal Sample Sizes

In the preceding section, the combined tests were each based on np observations; n from each $T_n^{(i)}$, $i=1,\ldots,p$. Now suppose n_i observations are used for the ith sequence, where

$$\lim_{n\to\infty} \frac{n_i}{np} = \lambda^{(i)} ,$$

$i=1,\ldots,p$, and $n_1+\ldots+n_p = n$. Then $\lambda^{(1)}+\ldots+\lambda^{(p)} = 1$, and

$$-\frac{2}{n_i} \cdot \log L_{n_i}^{(i)} \to c_i(\theta_i) \quad \text{as} \quad n_i \to \infty$$

with probability one $[\theta_i]$, so

$$\frac{-2}{n} \log L_{n_i}^{(i)} = \frac{n_i}{n} \cdot \frac{-2}{n_i} \log L_{n_i}^{(i)}$$

$$\to p\lambda^{(i)} c_i(\theta_i) .$$

With the three combined tests defined as in the previous sections, that is, with

$$T_n = \sqrt{-2 \log L_n^{(1)} \ldots L_n^{(p)}} ,$$

$$T_n^* = \frac{-2}{\sqrt{n}} \log \max_i L_n^{(i)} ,$$

and

$$\tilde{T}_n = \frac{-2}{\sqrt{n}} \log \min_i L_n^{(i)} ,$$

it follows that

$$\frac{T_n}{\sqrt{n}} = \sqrt{\frac{-2}{n} \log L_n^{(1)} + \dots + \frac{-2}{n} \log L_n^{(p)}}$$

$$\to \sqrt{p\lambda_1 c_1(\theta_1) + \dots + p\lambda_p c_p(\theta_p)}$$

$$= b(\theta) ,$$

$$\frac{T_n^*}{\sqrt{n}} = \min_i \frac{-2}{n} \log L_n^{(i)}$$

$$\to \min_i p\lambda_i c_i(\theta_i)$$

$$= b^*(\theta)$$

and

$$\frac{\tilde{T}_n}{\sqrt{n}} = \max_i \frac{-2}{n} \log L_n^{(i)}$$

$$= \min_i p\lambda_i c_i(\theta_i)$$

$$= \tilde{b}(\theta) .$$

Now, for the null θ , the distribution of the levels does not depend on the sample size, so

$$\frac{-1}{n} \log P_0\{T_n > \sqrt{n}t\} \to \frac{t^2}{2}$$

$$= f(t) ,$$

$$\frac{-1}{n} \log P_0\{T_n^* > \sqrt{n}t\} \to \frac{pt}{2}$$

$$= f^*(t) ,$$

and $$-\frac{1}{n} \log P_0\{\tilde{T}_n \to \sqrt{n}t\} \to \frac{t}{2}$$

$$= \tilde{f}(t) \; .$$

Hence, $$c(\theta) = 2f(b(\theta))$$

$$= p\lambda_1 c_1(\theta_1) + \; \ldots \; + p\lambda_p c_p(\theta_p) \; ,$$

$$c^*(\theta) = 2f^*(b^*(\theta))$$

$$= p \min_i p\lambda_i c_i(\theta_i) \; ,$$

and $$\tilde{c}(\theta) = 2\tilde{f}(\tilde{b}(\theta))$$

$$= \max_i p\lambda_i c_i(\theta_i) \; .$$

Thus, the slopes of the combined tests for unequal sample sizes follow the same form as for equal sample sizes with the individual slopes weighted in the same ratio as the sample sizes.

Comparison of the Three Methods of Combining

Observe that all three sequences of tests, $\{T_n\}$, $\{T_n^*\}$, and $\{\tilde{T}_n\}$, are calculated from the levels of the individual tests, based upon the product, the maximum, and the minimum of the levels, respectively. Thus, for $p=2$, rejection regions for the tests would be bounded by curves as illustrated in Figure 1.

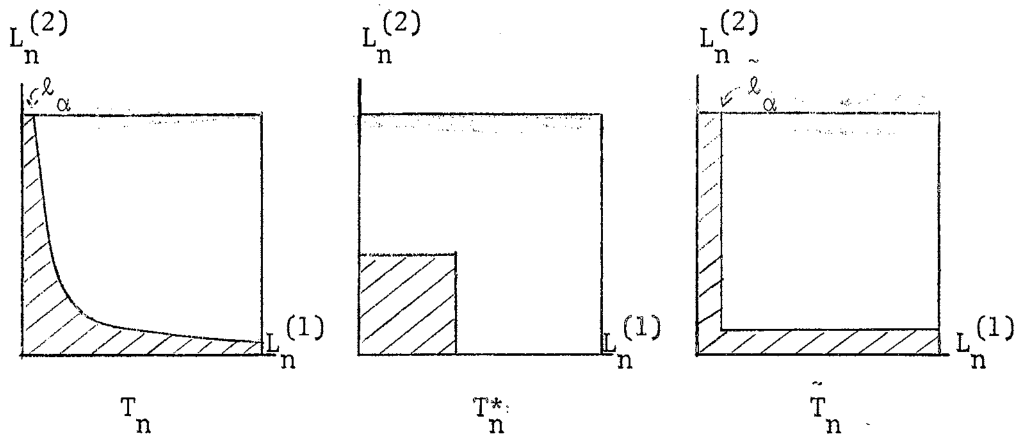

Figure 1.

Recall that the null hypothesis is $H: (\theta_1,\ldots,\theta_p)\varepsilon\Theta_0^{(1)}x\ldots x\Theta_0^{(p)}$, and the alternative is $A: (\theta_1,\ldots,\theta_p)\varepsilon\Theta^{(1)}x\ldots x\Theta^{(p)}-\Theta_0^{(1)}x\ldots x\Theta_0^{(p)}$. Thus H is false if any of the individual $H^{(i)}$ are false. If, say, for $p=2$, $H_0^{(1)}$ is false and $H_0^{(2)}$ is true, then one would expect small values for $L_n^{(1)}$ but not necessarily so for $L_n^{(2)}$, and for this case Figure 1 indicates that $\{T_n\}$ or $\{\tilde{T}_n\}$ would be more liekly to reject H than would $\{T_n^*\}$. The superiority of $\{T_n\}$ and $\{\tilde{T}_n\}$ over $\{T_n^*\}$ for this type of situation is also reflected in the exact slopes, since $\theta_1\varepsilon\Theta^{(1)}-\Theta_0^{(1)}$ and $\theta_2\varepsilon\Theta_0^{(2)}$ yields $c_1(\theta_1) > 0$ and $c_2(\theta_2) = 0$, and hence $c(\theta) = c_1(\theta_1)$, $c^*(\theta) = 0$, and $\tilde{c}(\theta) = c_1(\theta_1)$. On the other hand, if $H^{(1)}$ and $H^{(2)}$ are both false, then $\{T_n^*\}$ would appear more likely to yield rejection of H_0 than in the previous case when only $H^{(1)}$ were false. This is also reflected in the slopes, especially for θ_1 and θ_2 such that $c_1(\theta_1)$ and $c_2(\theta_2)$ are nearly equal. For then $c(\theta) = c_1(\theta_1) + c_2(\theta_2)$ is approximately the same as $\tilde{c}(\theta) = 2\min\{c_1(\theta_1), c_2(\theta_2)\}$, and $c^*(\theta) = \max\{c_1(\theta_1), c_2(\theta_2)\}$ is only about half as large as either $c(\theta)$ or $\tilde{c}(\theta)$. Figure 2 is useful for determining the relative sizes of the three exact slopes $c(\theta)$, $c^*(\theta)$ and $\tilde{c}(\theta)$.

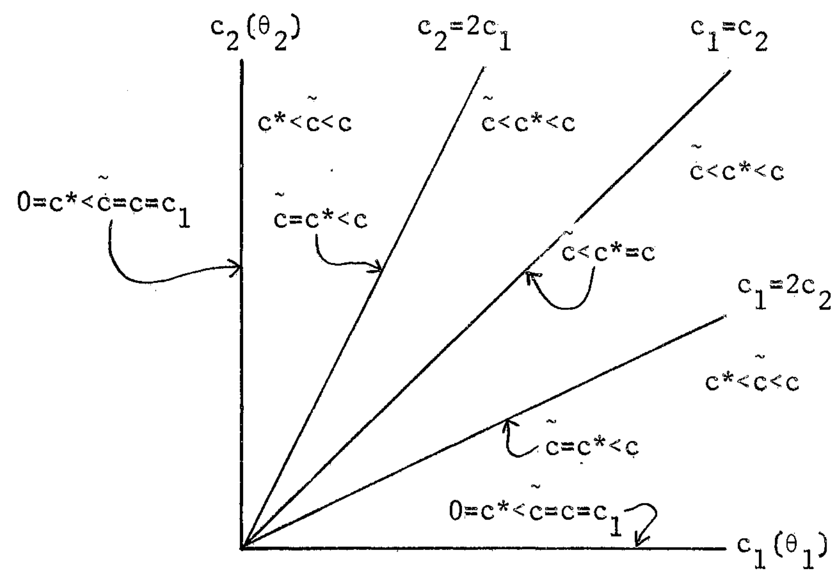

Figure 2.

As an application of comparing the combining procedures, consider an example given by Naik (11).

Example 4.1: Let $X_1^{(i)}, \ldots, X_n^{(i)}$ be a random sample of size n from a normal distribution with mean θ_i and variance one, $i=1,2$, and consider $H^{(i)}: \theta_i \leq 0$ and $A^{(i)}: \theta_i > 0$, $i=1,2$. Then the combined hypotheses are $H: \theta_1 \leq 0$ and $\theta_2 \leq 0$ versus $A: \theta_1 > 0$ or $\theta_2 > 0$, so $\theta=(\theta_1, \theta_2)$, $\Theta=(-\infty, \infty)x(-\infty, \infty)$, and $\Theta_0=(-\infty, 0]x(-\infty, 0]$. Take $T_n^{(i)}$ to be based on the sample mean of the ith sample; that is, $T_n^{(i)} = \sqrt{n}\,\bar{X}^{(i)}$. Now the distribution of $T_n^{(i)}$ depends on θ_i even for $\theta_i \varepsilon \Theta_0^{(i)}$, so let the level attained by $T_n^{(i)}$ be defined by

$$L_n^{(i)}(t) = \sup_{\theta_i \varepsilon \Theta_0^{(i)}} [1-F_n(t,\theta)] .$$

(This is the definition of $L_n(t)$ taken by R. R. Bahadur when he wishes to consider tests whose distribution may depend on $\theta \varepsilon \Theta_0$, e.g. in (2).) In this example,

$$\sup_{\theta_i \varepsilon \Theta_0^{(i)}} [1-F_n(t,\theta)] = \sup_{\theta_i \leq 0} [1 - \int_{-\infty}^{t} N(x;\sqrt{n}\theta_i,1)dx]$$

$$= \sup_{\theta_i \leq 0} [1 - \int_{-\infty}^{t-\sqrt{n}\theta_i} N(x;0,1)dx]$$

$$= 1 - \inf_{\theta_i \leq 0} \int_{-\infty}^{t-\sqrt{n}\theta_i} N(x;0,1)dx$$

$$= 1 - \int_{-\infty}^{t} N(x;0,1)dx$$

$$= 1 - \Phi(t) ,$$

so $L_n(t) = 1-\Phi(t)$, as in the case of $\Theta_0^{(i)} = \{0\}$. Hence, as was seen in Example 2.1, $c_i(\theta_i) = \theta_i^2$ for $\theta_i > 0$.

Now, since $\Theta-\Theta_0 = (-\infty, \infty)x(-\infty, \infty) - (-\infty,0]x(-\infty,0]$, there are θ's in $\Theta-\Theta_0$ for which θ_1 or θ_2 (but not both) may belong to $\Theta_0^{(1)}$ or $\Theta_0^{(2)}$, respectively. Thus, in order to calculate $c(\theta)$ on

$\Theta - \Theta_0$, it is necessary to have $c_i(\theta_i)$ defined on $\Theta_0^{(i)}$, $i=1,2$; that is, for $\theta_i \leq 0$. For $\theta_i = 0$, it was seen in Chapter III that

$$\frac{-2}{n} \log L_n(T_n) \to 0$$

with probability one $[\theta=0]$, so $c_i(0) = 0$. Also, for $\theta_i < 0$,

$$T_n = \sqrt{n}\,\overline{X} \to -\infty$$

with probability one $[\theta]$ and thus

$$-\frac{2}{n} \log L_n(T_n) = -\frac{2}{n} \log\,[1-\Phi(T_n)]$$

$$\to 0$$

with probability one $[\theta]$. Therefore, $c_i(\theta_i) = \theta_i^2$, for $\theta_i > 0$, and $c_i(\theta_i) = 0$ for $\theta_i \leq 0$. It then follows that the slopes for the combined tests are given as illustrated in Figure 3.

Naik (11) has made "equal probability" comparisons of Fisher's Method $(\{T_n\})$ and the method based on minimum level $(\{\tilde{T}_n\})$, which he refers to as the Union-Intersection Principle . This is done by bounding θ_i away from zero in the alternative parameter space by a distance $\delta > 0$. This subset E_0 of the alternative space is the region located above and to the right of the dotted lines in Figure 3. Then for each method of combining, an α^* is found such that the test has type one error probability of α^* (at $(\theta_1, \theta_2) = (0,0)$) and also has maximum type two error probability of α^* , where the

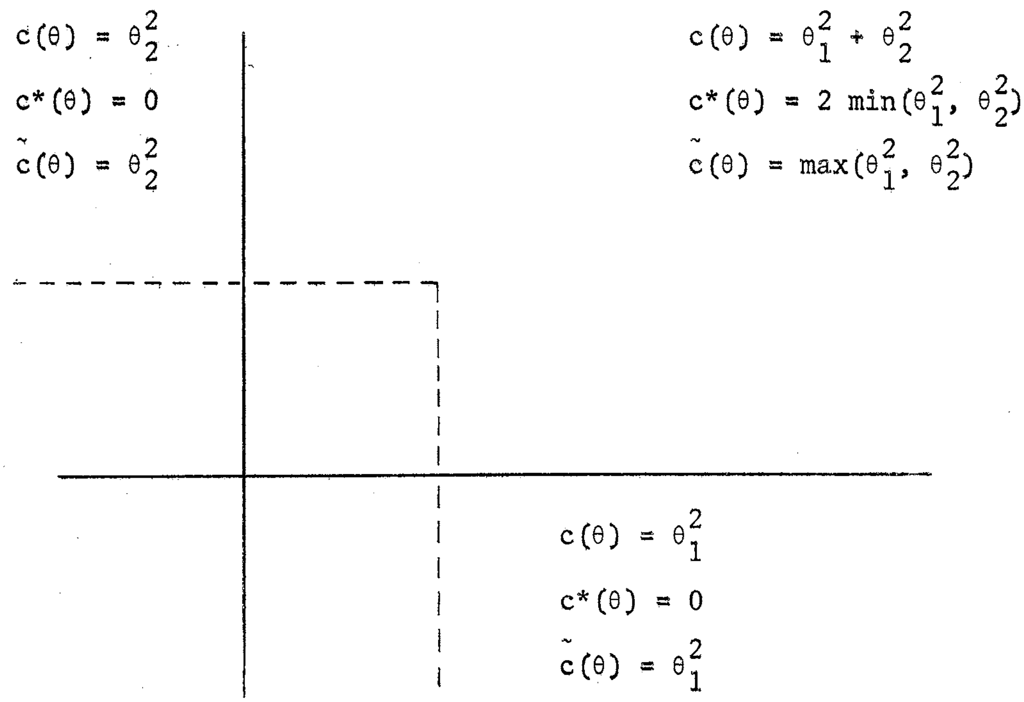

Figure 3.

maximization is taken over $(\theta_1, \theta_2) \varepsilon E_0$. This maximum occurs for one of the θ_i equal to δ and the other equal to minus infinity (See Theorem A.2 of the Appendix). Then the method of combining which yields the smaller α^* is judged superior to the method which yields the larger α^*. A table in (11) calculated for various values of $\sqrt{n}\,\delta$ gives uniformly smaller α^* values for the Union-Intersection Principle (U-IP) than for Fisher's method (FM), indicating superiority of the U-IP over FM. In the Appendix it is shown that for one of the θ_i equal to δ and the other equal to $-\infty$ and for all sizes α, the U-IP yields a more powerful test than does FM. That is,

$$1-\tilde{\beta}(\alpha) > 1-\beta(\alpha) \, , \qquad 0 < \alpha < 1 \, ,$$

where $1-\tilde{\beta}(\alpha)$ and $1-\beta(\alpha)$ are the powers of the combined tests of size α using the U-IP and FM, respectively. The quantities $\tilde{\alpha}^*$ and α^* are found by solving the equations

$$\tilde{\beta}(\tilde{\alpha}^*) = \tilde{\alpha}^*$$

and

$$\beta(\alpha^*) = \alpha^* \, .$$

Since $1-\beta(\alpha)$ is an increasing function of α, it follows that $\tilde{\alpha}^* < \alpha^*$, because $\tilde{\alpha}^* \geq \alpha^*$ implies

$$\begin{aligned} 1-\beta(\tilde{\alpha}^*) &\geq 1-\beta(\alpha^*) \\ &= 1-\alpha^* \\ &\geq 1-\tilde{\alpha}^* \\ &= 1-\tilde{\beta}(\tilde{\alpha}^*) \, , \end{aligned}$$

which contradicts $1-\beta(\alpha) < 1-\tilde{\beta}(\alpha)$ for all α. Thus, in fact, according to the equal probability criterion, the U-IP is superior to FM.

Bahadur efficiency does not distinguish between the U-IP and FM of combining the tests of this example at precisely the same parameter values employed in making the equal probability comparison (i.e., one of θ_i equal to δ and the other equal to $-\infty$), for there $c(\theta) = \tilde{c}(\theta) = \delta^2$. However, in the region where Bahadur efficiency does distinguish, that is, in the first quadrant, it judges FM superior to U-IP because $c(\theta) = \theta_1^2 + \theta_2^2 > \max(\theta_1^2, \theta_2^2) = \tilde{c}(\theta)$. The relation between Bahadur efficiency and power might lead one to conjecture that for $\theta = (\theta_1, \theta_2)$ in the first quadrant, FM gives a more powerful combined test than the U-IP. Theorem 2.2 implies this is true for at least some α .

Optimality of Fisher's Method

It is clear that the sequence $\{T_n\}$ obtained by Fisher's method always yields a larger exact slope than the sequence $\{T_n^*\}$ based on maximum level or the sequence $\{\tilde{T}_n\}$ based on minimum level. It is shown in the following theorem that if $\{T_n^{(i)}\}$ has maximum slope for testing $H^{(i)}: \theta_i \varepsilon \Theta_0^{(i)}$, i=1,...,p , then Fisher's method is the optimal method of combining the $\{T_n^{(i)}\}$ in order to test $H_0: (\theta_1,\ldots,\theta_p)\varepsilon\Theta_0^{(1)} x \ldots x \Theta_0^{(p)}$.

<u>Theorem 4.1</u>: Suppose $\{T_n^{(i)}\}$ has maximum slope for testing $H^{(i)}: \theta_i \varepsilon \Theta_0^{(i)}$, i=1,...,p . If $c'(\theta)$ is the exact slope of any sequence of tests $\{T_n'\}$ obtained by combining the $\{T_n^{(i)}\}$ for testing $H : (\theta_1,\ldots,\theta_p) \varepsilon\ \Theta_0^{(1)} x \ldots x\ \Theta_0^{(p)}$, then $c(\theta) \geq c'(\theta)$, where $c(\theta)$ is the exact slope of the sequence $\{T_n\}$ obtained by Fisher's method. In fact, $c(\theta)$ is the maximum slope of all sequences of tests of H , combined or not.

Proof: Since $\{T_n^{(i)}\}$ is optimal for $H^{(i)}$, it follows that $c_i(\theta_i) = 2 \inf_{\theta_{i0} \varepsilon \Theta_0^{(i)}} K(\theta_i, \theta_{i0})$ by Theorem 1.2. View the data as

$$x_1 = (x_1^{(1)}, \ldots, x_1^{(p)})$$
$$\vdots$$
$$x_n = (x_n^{(1)}, \ldots, x_n^{(p)}) .$$

Then the probability density function $h_\theta(x)$ is

$$h_\theta(x) = f^{(1)}_{\theta_1}(x^{(1)}) f^{(2)}_{\theta_2}(x^{(2)}) \ldots f^{(p)}_{\theta_p}(x^{(p)}) ,$$

where $f^{(i)}_{\theta_i}(x^{(i)})$ is the probability density function of the ith random variable $X^{(i)}$. Hence, by Theorem 1.2, the maximum slope of a sequence $\{T'_n\}$ based on $\{x_n\} = \{(x^{(1)}_n, \ldots, x^{(p)}_n)\}$ is given by

$$\begin{aligned}
2 \inf_{\theta_0 \varepsilon \Theta_0} K(\theta, \theta_0) &= 2 \inf_{\theta_0 \varepsilon \Theta_0} E_\theta\left[\log \frac{h_\theta(X)}{h_{\theta_0}(X)}\right] \\
&= 2 \inf_{\theta_0 \varepsilon \Theta_0} E_\theta\left[\log \frac{\prod_i f^{(i)}_{\theta_i}(X^{(i)})}{\prod_i f^{(i)}_{\theta_{i0}}(X^{(i)})}\right] \\
&= 2 \inf_{\theta_0 \varepsilon \Theta_0} E_\theta\left[\sum_i \log \frac{f^{(i)}_{\theta_i}(X^{(i)})}{f^{(i)}_{\theta_{i0}}(X^{(i)})}\right] \\
&= 2 \inf_{\theta_0 \varepsilon \Theta_0} \sum_i E_\theta \log \frac{f^{(i)}_{\theta_i}(X^{(i)})}{f^{(i)}_{\theta_{i0}}(X^{(i)})} \\
&= 2 \inf_{\theta_0 \varepsilon \Theta_0} \sum_i K^{(i)}(\theta_i, \theta_{i0}) \\
&= \sum_i 2 \inf_{\theta_{i0} \varepsilon \Theta^{(i)}_0} K^{(i)}(\theta_i, \theta_{i0}) \\
&= \sum_i c_i(\theta_i) \\
&= c(\theta) ,
\end{aligned}$$

which completes the proof.

CHAPTER V

EXTENSIONS

In Chapter II, an example was given to show that Bahadur efficiency is not always in the same agreement with power as is stated in (1), (7). This raises the question of whether in general a sequence constructed as $\{T_n^{(1)}\}$ of Example 2.1 was constructed will result in $\phi_{12} = 1$. Also, in Chapter II examples of various behavior of the ratio of levels are given. However, some situations are not illustrated; in particular, examples of the ratio converging to finite nonzero constants would be interesting. An example with approximate efficiency equal to one and a limit of ratios of approximate levels equal to zero is given, but the author has been unable to construct an example of the analogous situation involving exact slopes.

There are undoubtedly aspects of information theory other than those in Chapter III which have analogies in terms of exact slopes. For instance, theorems in information theory concerning the concept of sufficiency could lead to a search for similar relations between exact slopes and sufficiency.

It seems quite desirable to extend the theory of Chapter IV to combining tests which are not necessarily independent. Following Fisher's method, one might base a combined test on

$$P_0\{T_n^{(1)} > T_n^{(1)}(s),\ T_n^{(2)} > T_n^{(2)}(s)\} =$$

$$= P_0\{T_n^{(1)} > T_n^{(1)}(s)\} \cdot P_0\{T_n^{(2)} > T_n^{(2)}(s) | T_n^{(1)} > T_n^{(1)}\}$$

The difficulty with this approach lies in the fact that $P_0^{(1)}\{T_n^{(2)} > T_n^{(2)}(s) | T_n^{(1)} > T_n^{(1)}(s)\}$ does not appear to be uniformly distributed for $\theta \varepsilon \Theta_0$, and hence the quantity

$$-2 \log P_0\{T_n^{(1)} > T_n^{(1)}(s),\ T_n^{(2)} > T_n^{(2)}(s)\}$$

$$= -2 \log P_0\{T_n^{(1)} > T_n^{(1)}(s)\} - 2 \log P_0\{T_n^{(1)} > T_n^{(1)}(s) | T_n^{(2)} > T_n^{(2)}(s)\}$$

does not have a chi-square distribution, as was the case with independent tests.

A theory of combining dependent tests would appear to have applications in the investigation of the role of sufficiency with regard to exact slopes. For instance, if it happens that

$$c(\theta) = c_1(\theta) + c_{2|1}(\theta)$$

where $c(\theta)$ is the slope of the combined test, $c_1(\theta)$ is the slope of $\{T_n^{(1)}\}$ and $c_{2|1}(\theta)$ is the slope of $\{T_n^{(2)}\}$ given $\{T_n^{(1)}\}$, then it would follow immediately that in searching for tests with maximum slope, one may limit the search to functions of a sufficient statistic.

A more practical application would be in measuring how much is gained in combining two tests, neither of which have maximum slope. An example of this situation is combining a sign test and a Kolmogorov-Smirnov type statistic for testing about a location parameter of a given family of distributions.

The relation between the equal probability criterion and Bahadur efficiency deserves further investigation. Conditions under which the two criteria agree would be desirable. The rate of convergence of α_n^* should also be studied, where, for fixed $\theta \varepsilon \Theta_0 - \Theta$, $\beta_n(\alpha_n^*, \theta) = \alpha_n^*$. It is shown in (2) that $-\frac{2}{n} \log \alpha_n \to c(\theta)$, where for fixed $\theta \varepsilon \Theta - \Theta_0$ and fixed β, $0 < \beta < 1$, $\beta_n(\alpha_n, \theta) = \beta$. The geometric interpretations of $\{\alpha_n\}$ and $\{\alpha_n^*\}$ are given in Figure 4, where the curves are those of $\beta = \beta_n(\alpha, \theta)$, n=1,2,...

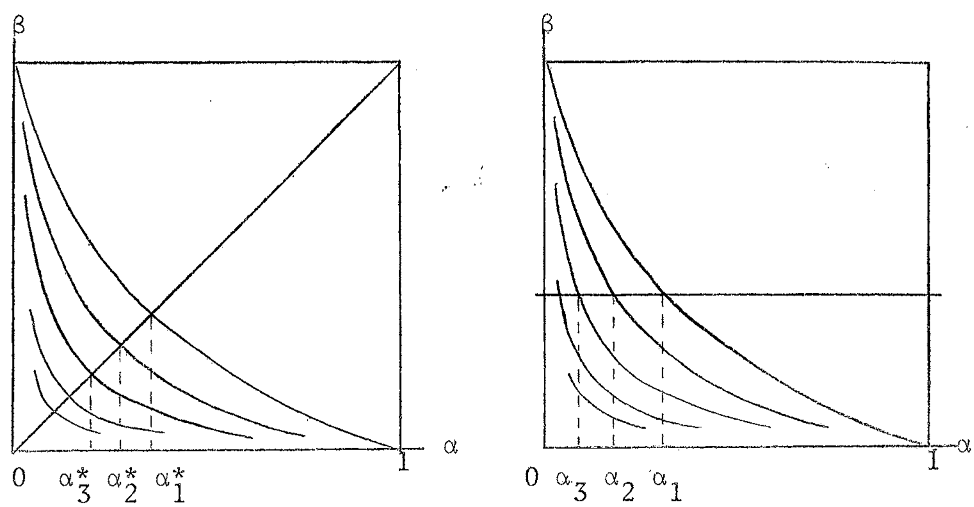

Figure 4

Two examples are now given, one in which the rate of convergence of α_n^* to zero is the same as the rate of convergence of α_n to zero, and one in which the two rates are much different.

Example 5.1: Let X_i , i=1,2,... be independent and identically distributed random variables with distribution given by $P_\theta\{X_i > x\} = e^{-(x-\theta)}$ for $x \geq \theta$ and $P_\theta\{X_i > x\} = 0$ for $x < \theta$. Define $T_n = \sqrt{n}\, X_{1n}$, where $X_{1n} = \min_{1\leq i\leq n} \{X_i\}$. Then $\frac{T_n}{\sqrt{n}} = X_{1n} \to \theta$ with probability one [θ]. Now

$$P_\theta\{X_{1n} > x\} = P_\theta\{X_i > x\ ,\ i=1,\dots,n\}$$
$$= e^{-n(x-\theta)}\ ,$$

hence

$$-\frac{1}{n} \log P_0\{T_n > \sqrt{n}t\} = -\frac{1}{n} \log P_0\{X_{1n} > t\}$$
$$= -\frac{1}{n} \log e^{-nt}$$
$$= t\ ,$$

So by Theorem 1.1, $c(\theta) = 2\theta$, and hence $-\frac{2}{n} \log \alpha_n \to 2\theta$, where $\beta_n(\alpha_n, \theta) = \beta$.

Let the critical regions of the equal probability test based on $\{T_n\}$ be given by $(\sqrt{n}t_n, \infty)$, n=1,2,... . That is, $\{t_n\}$ is a sequence of positive numbers such that

$$P_0\{T_n > \sqrt{n}t_n\} = \alpha_n^* = \beta_n(\alpha_n^*,\theta) = P_\theta\{T_n < \sqrt{n}t_n\}\ .$$

Then

$$e^{-nt_n} = 1 - e^{-n(t_n-\theta)}\ ,$$

so

$$e^{-nt_n}[1+e^{n\theta}] = 1\ .$$

Therefore

$$\alpha_n^* = e^{-nt_n}$$

$$= \frac{1}{1+e^{n\theta}}$$

$$= e^{-n\theta} - e^{-2n\theta} + e^{-3n\theta} - e^{-4n\theta} + \ldots$$

$$= e^{-n\theta}[1+o(1)] \quad \text{as} \quad n\to\infty \ .$$

It follows that

$$-\frac{2}{n}\log \alpha_n^* = -\frac{2}{n}\log e^{-n\theta}[1+o(1)]$$

$$= 2\theta - \frac{2}{n}\log [1+o(1)]$$

$$\to 2\theta \quad \text{as} \quad n\to\infty \ ,$$

and thus the sequences $\{\alpha_n\}$ and $\{\alpha_n^*\}$ converge to zero at the same exponential rate.

Example 5.2: Let $X_1, X_2, \ldots$ be independently and identically distributed normal random variables with mean μ and variance one. Define $T_n = \sqrt{n}\,\bar{X}$. It was seen in Example 2.1 that $c(\mu) = \mu^2$, and thus $-\frac{2}{n}\log \alpha_n \to \mu^2$, where $\beta_n(\alpha_n, \mu) = \beta$. Now for fixed $\mu > 0$, take t_n such that

$$\alpha_n^* = P_0\{T_n > \sqrt{n}t_n\} = P_\mu\{T_n < \sqrt{n}t_n\} = \beta_n(\alpha_n^*, \mu) \ .$$

Now

$$P_0\{T_n > \sqrt{n}t_n\} = \int_{\sqrt{n}t_n}^{\infty} n(x;0,1)dx$$

and

$$P_\mu\{T_n < \sqrt{n}t_n\} = \int_{-\infty}^{\sqrt{n}t_n} n(x;\sqrt{n}\,\mu,1)dx$$

$$= \int_{-\infty}^{\sqrt{n}t_n - \sqrt{n}\mu} n(x;0,1)dx$$

$$= \int_{\sqrt{n}(\mu-t_n)}^{\infty} n(x;0,1)dx ,$$

so

$$\sqrt{n}(\mu-t_n) = \sqrt{n}t_n$$

and hence

$$t_n = \frac{\mu}{2}$$

Thus

$$\frac{-2}{n}\log \alpha_n^* = \frac{-2}{n}\log P_0\{T_n > \sqrt{n}\,\frac{\mu}{2}\}$$

$$\rightarrow (\frac{\mu}{2})^2$$

$$= \frac{\mu^2}{4}$$

by results of Example 2.1.

Figure 5 shows the areas which give the type one and type two error probabilities for Examples 5.1 and 5.2. The curves represent the density functions of $\frac{T_n}{\sqrt{n}} = X_{1n}$ of Example 5.1 and $\frac{T_n}{\sqrt{n}} = \overline{x}$ of Example 5.2.

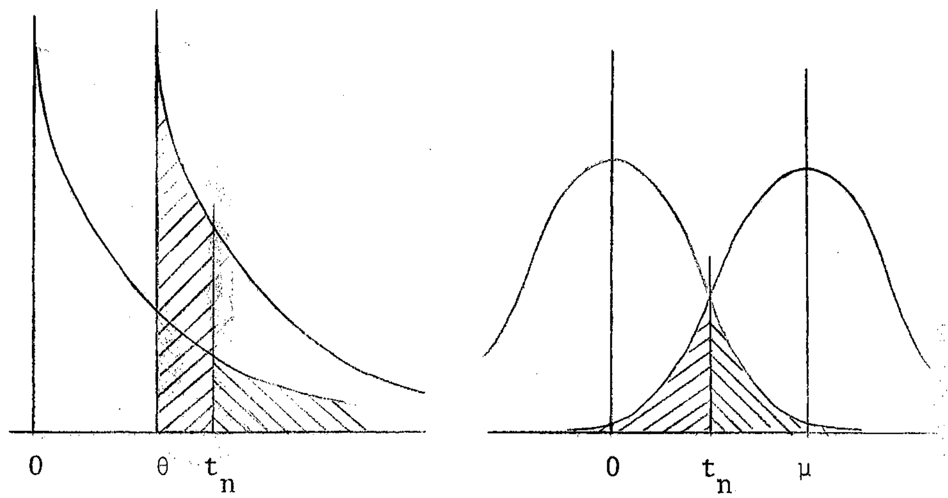

Figure 5

Notice that it is the highly skewed distribution of Example 5.1 for which the exponential rates of convergence of α_n and α_n^* are the same, whereas the rates are quite different for the symmetrical distribution of Example 5.2. This is because the numbers t_n of Example 5.1 tend to be closer to the alternative parameter θ than to the null parameter value of zero, and in Example 5.2, the t_n quantities are midway between the null parameter value of zero and the alternative value of μ. It would thus appear that, at least for hypotheses concerning location parameters, the more the distribution is skewed to the right, the more nearly the same are the rate of convergence of the equal probability error and the rate of convergence of the type one error with fixed type two error.

CHAPTER VI

SUMMARY

This investigation dealt with the role of Bahadur efficiency in certain apsects of the foundations of statistical inference. In Chapter I, Bahadur efficiency was presented, and examples were given illustrating its application.

Chapter II was concerned with the relation between Bahadur efficiency and a criterion called domination, which has to do with power, and the relation between Bahadur efficiency and the convergence of the ratio of significance levels. An example was given which shows that Bahadur efficiency and power are not quite as closely related as previously thought. It was shown that if $\phi_{12} < 1$, then $\frac{L_n^{(2)}}{L_n^{(1)}}$ converges to zero, and examples were given showing various behaviors of the ratio of levels when $\phi_{12} = 1$.

Analogies were drawn in Chapter III between properties of information functions and exact slopes. The analogies held for properties of non-negativity, invariance, lack of information in statistics whose distribution is independent of θ, inability to increase information by data manipulation, and the additivity of information in independent observations.

In Chapter IV, it was shown that the exact slopes of the combined tests are $\sum_i c_i(\theta_i)$, $p \min_i c_i(\theta_i)$ and $\max_i c_i(\theta_i)$, where the combining is based on the product (Fisher method), the maximum level, and the minimum level, respectively. It was shown that Bahadur efficiency is not in complete agreement with the equal probability criterion, when it comes to choosing between methods of combining.

A SELECTED BIBLIOGRAPHY

(1) Bahadur, R. R. "Stochastic Comparisons of Tests." Annals of Mathematical Statistics. Vol. 31. (1960) 276-295.

(2) Bahadur, R. R. "Rates of Convergence of Estimates and Test Statistics." Annals of Mathematical Statistics. Vol. 38. (1967) 303-324.

(3) Bahadur, R. R. "Asymptotic Efficiency of Tests and Estimates." Sankhya. Vol. 22. (1960) 229-252.

(4) Bahadur, R. R. Conference on Limit Theorems in Statistics. Department of Statistics, Florida State University. (mimeographed lecture notes, 1969).

(5) Feller, W. An Introduction to Probability Theory and Its Applications, Vol. 1, 2 nd ed. John Wiley and Sons, Inc., New York, N.Y. (1957).

(6) Fisher, R. A. Statistical Methods for Research Workers, 11th ed. Oliver and Boyd, London (1950).

(7) Gleser, L. J. "On a Measure of Test Efficiency Proposed by R. R. Bahadur." Annals of Mathematical Statistics. Vol. 35. (1964) 1537-1544.

(8) Hoardley, A. Bruce. "On the Probability of Large Deviations of Functions of Several Emperical cdf's." Annals of Mathematical Statistics. Vol. 38. (1967) 360-382.

(9) Hogg, Robert V. and Craig, Allen T. Introduction to Mathematical Statistics, 2nd ed. The Macmillan Company, New York, N.Y. (1965).

(10) Moore, Billy J. "On a Theory of Significance Testing." Unpublished Ph.D. Thesis (1969). Oklahoma State University.

(11) Naik, Umesh D. "The Equal Probability Test and Its Applications to Some Simultaneous Inference Problems." Journal of the American Statistical Association. Vol. 64. (1969) 986-998.

(12) Papaioannou, Takis. <u>On Statistical Information Theory and Related Measures of Information</u>. Statistical Laboratory, Iowa State University. (mineographed lecture notes, 1969).

(13) Savage, I. R. "Nonparametric Statistics: A Personal Review." <u>Sankhya</u>. Vol. 31. (1969) 107-144.

(14) Seivers, Gerald L. "On the Probability of Large Deviations and Exact Slopes." <u>Annals of Mathematical Statistics</u>. Vol. 40. (1969) 1908-1922.

APPENDIX

The following example illustrates the existence of unbiased test statistics for the hypothesis H: $\theta=\theta_0$ versus A: $\theta\neq\theta_0$ which are non-unimodal functions of a statistic X , whose density has strict monotone likelihood ratio in x . See (10) for a complete discussion.

Example A.1: Let $X_1, \ldots X_n$ denote a random sample from a normal distribution with mean μ and variance one, and consider the hypotheses H: $\mu=0$ and A: $\mu\neq 0$. Define

$$\tilde{T}_n = |\sqrt{n}\,\overline{X}|^{-1} e^{\frac{n\overline{X}^2}{2}} .$$

It is clear that $\tilde{T}_n$ is a bimodal function of $\overline{x}$.

Let t_α be such that $P_0\{\tilde{T}_n > t_\alpha\} = \alpha$. (Notice that t_α does not depend on n , since the distribution of $\sqrt{n}\,\overline{X}$, and hence of T_n , is the same for all n when $\mu=0$.) The power of $\tilde{T}_n$ is given by

$$1-\beta_n(\alpha, \mu) = P_\mu\{\tilde{T}_n > t_\alpha\}$$

$$= P_\mu\{\sqrt{n}\,\overline{X} < -a_\alpha\}+P_\mu\{-b_\alpha<\sqrt{n}\overline{X}<b_\alpha\}+P_\mu\{\sqrt{n}\overline{X} > a_\alpha\} ,$$

where $0 < b_\alpha < a_\alpha$ and $t_\alpha = a_\alpha^{-1} e^{\frac{a_\alpha^2}{2}} = b_\alpha^{-1} e^{\frac{b_\alpha^2}{2}}$.

This gives

$$1-\beta_n(\alpha, \mu) = \int_{-\infty}^{-a_\alpha} n(x;\sqrt{n}\mu,1)dx + \int_{-b_\alpha}^{b_\alpha} n(x;\sqrt{n}\mu,1)dx + \int_{a_\alpha}^{\infty} n(x;\sqrt{n}\mu,1)dx .$$

Now, for constants c_1 and c_2,

$$\int_{c_1}^{c_2} n(x;\sqrt{n}\mu,1)dx = \int_{c_1-\sqrt{n}\mu}^{c_2-\sqrt{n}\mu} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx ,$$

so

$$\frac{\partial}{\partial\mu} \int_{c_1}^{c_2} n(x;\sqrt{n}\mu,1)cx = \frac{\sqrt{n}}{\sqrt{2\pi}} \left[e^{-(c_1-\sqrt{n}\mu)^2/2} - e^{-(c_2-\sqrt{n}\mu^2)/2} \right].$$

Thus

$$\frac{\partial}{\partial\mu}[1-\beta_n(\alpha, \mu)] = \sqrt{\frac{n}{2\pi}} [(-e^{-(-a_\alpha-\sqrt{n}\mu)^2/2} + (e^{-(-b_\alpha-\sqrt{n}\mu)^2/2}$$

$$-e^{-(b_\alpha-\sqrt{n}\mu)^2/2}) + (e^{-(a_\alpha-\sqrt{n}\mu)^2/2})]$$

$$= \sqrt{\frac{n}{2\pi}}\ e^{-\frac{n\mu^2}{2}} [e^{-a_\alpha^2/2} (e^{a_\alpha\sqrt{n}\mu} - e^{-a_\alpha\sqrt{n}\mu})$$

$$- e^{-b_\alpha^2/2} (e^{b_\alpha\sqrt{n}\mu} - e^{-b_\alpha\sqrt{n}\mu})]$$

$$= \sqrt{\frac{n}{2\pi}}\ e^{-\frac{n\mu^2}{2}}\ h_n(\alpha, \mu) ,$$

where $h_n(\alpha, \mu) = e^{-a_\alpha^2/2} (e^{a_\alpha\sqrt{n}\mu} - e^{-a_\alpha\sqrt{n}\mu}) - e^{-b_\alpha^2/2} (e^{b_\alpha\sqrt{n}\mu} - e^{-b_\alpha\sqrt{n}\mu})$.

It is clear that $\frac{\partial}{\partial\mu}[1-\beta_n(\alpha, \mu)] = 0$ if and only if $h_n(\alpha, \mu) = 0$.

Now

$$\frac{\partial}{\partial\mu} h_n(\alpha, \mu) = \sqrt{n}\ a_\alpha\ e^{-a_\alpha^2/2} (e^{a_\alpha\sqrt{n}\mu} + e^{-a_\alpha\sqrt{n}\mu})$$

$$- \sqrt{n} b_\alpha\ e^{-b_\alpha^2/2} (e^{b_\alpha\sqrt{n}\mu} + e^{-b_\alpha\sqrt{n}\mu})$$

$$= 2\sqrt{n}\, t_\alpha^{-1}[\cosh(a_\alpha\sqrt{n}\mu) - \cosh(b_\alpha\sqrt{n}\mu)]$$

$$> 0 \quad \text{for} \quad \mu > 0 \quad \text{since} \quad 0 < b_\alpha < a_\alpha \,.$$

Hence $h_n(\alpha, \mu) = \int_0^\mu [\frac{\partial}{\partial\mu} h_n(\alpha, \nu)]d\nu$

$$> 0 \quad \text{for} \quad \mu > 0 \,,$$

so $$\frac{\partial}{\partial\mu}[1-\beta_n(\alpha, \mu)] = \sqrt{\frac{n}{2\pi}}\; e^{-\frac{n\mu^2}{2}} \cdot h_n(\alpha, \mu)$$

$$> 0 \quad \text{for} \quad \mu > 0 \,.$$

Thus $1-\beta_n(\alpha, -\mu)$ is strictly increasing in μ for $\mu > 0$. Also, $1-\beta_n(\alpha, \mu) = 1-\beta_n(\alpha, -\mu)$, so $1-\beta_n(\alpha, \mu)$ is minimized for $\mu=0$, and therefore $\tilde{T}_n$ is unbiased.

It will now be shown that the statistic $|\sqrt{n}\overline{X}|^{-1}\, e^{\frac{n\overline{X}^2}{2}}$ has exact slope $c(\mu) = \mu^2$. Let $Y_n = n\overline{X}^2$, which has a chi-square distribution with one degree of freedom when $\mu=0$ for each n. In order to use Theorem 1.1, define

$$T_n^* = \sqrt{\log(Y_n^{-1/2}\, e^{Y_n/2})}$$

$$= \sqrt{\tfrac{1}{2}(Y_n - \log Y_n)} \,.$$

Obviously, T_n^* is equivalent to $\tilde{T}_n = |\sqrt{n}\overline{X}|^{-1}\, e^{n\overline{X}^2/2}$.

Now

$$\frac{T_n^*}{\sqrt{n}} = \sqrt{\frac{1}{2}\left(\frac{Y_n}{n} - \frac{\log Y_n}{n}\right)}$$

$$\to \sqrt{\tfrac{1}{2}(\mu^2 - 0)}$$

$$= \sqrt{\frac{\mu^2}{2}} ,$$

so, for i) of Theorem 1.1, $b(\mu) = \sqrt{\frac{\mu^2}{2}}$.

Also,

$$P_0\{T_n^* \geq \sqrt{n}\, \nu\} = P_0\{\sqrt{\tfrac{1}{2}(Y_n - \log Y_n)} \geq \sqrt{n}\nu\}$$

$$= P_0\{Y_n - \log Y_n \geq 2n\nu^2\}$$

$$= P_0\{Y_n \leq y_n'\} + P_0\{Y_n \geq y_n''\} ,$$

where y_n' and y_n'' are the two solutions to $y_n - \log y_n = 2n\nu^2$, and $0 < y_n' < y_n''$.

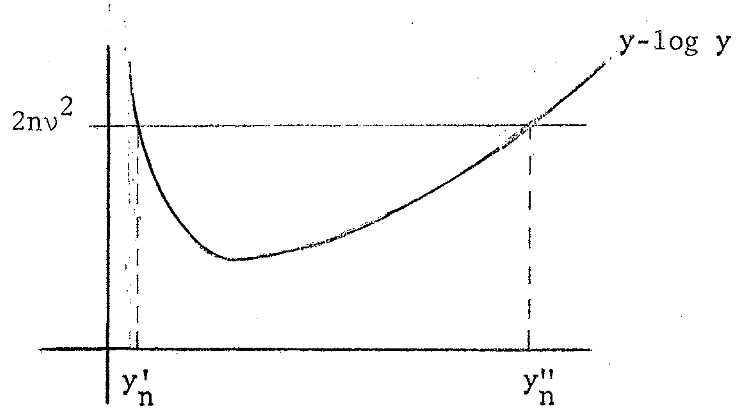

Figure 6

It is clear from Figure 6 that $y_n' \to 0$ and $y_n'' \to \infty$ as $n \to \infty$, so $\dfrac{\log y_n''}{y_n''} \to 0$ as $n \to \infty$. From the equation $y_n'' - \log y_n'' = 2n\nu^2$ it follows that $\dfrac{y_n''}{2n}[1 - \dfrac{\log y_n''}{y_n''}] = \nu^2$ for each n , and hence $\dfrac{y_n''}{2n} \to \nu^2$.

Thus

$$-\frac{1}{n}\log P_0\{Y_n > y_n''\} = -\frac{1}{n}\log P_0\{\sqrt{Y_n} > \sqrt{y_n''}\}$$

$$= -\frac{1}{n}\log e^{-\frac{1}{2}(\sqrt{y_n''})^2[1+o(1)]}$$

$$= \frac{1}{2n}\, y_n''[1+o(1)]$$

$$\to \nu^2 \quad \text{as} \quad n\to\infty \text{ .}$$

Since $y_n' - \log y_n' = 2n\nu^2$, it follows that

$$-\frac{1}{n}\log y_n' = \frac{1}{n}(2n\nu^2 - y_n')$$

$$\to 2\nu^2 \quad \text{as} \quad n\to\infty \text{ .}$$

Now

$$P_0\{Y_n \leq y_n'\} = \frac{1}{\Gamma(\frac{1}{2})2^{\frac{1}{2}}}\int_0^{y_n'} t^{-\frac{1}{2}} e^{-\frac{1}{2}}\, dt \text{ .}$$

But

$$e^{-\frac{y_n'}{2}}\, 2\sqrt{y_n'} = e^{-\frac{y_n'}{2}}\int_0^{y_n'} t^{-\frac{1}{2}}\, dt$$

$$< \int_0^{y_n'} t^{-\frac{1}{2}} e^{-\frac{t}{2}}\, dt$$

$$< \int_0^{y_n'} t^{-\frac{1}{2}}\, dt$$

$$= 2\sqrt{y_n'} \text{ ,}$$

and

$$-\frac{1}{n}\log e^{-\frac{y_n'}{2}} = \frac{y_n'}{2n}$$

$$\to 0 \ ,$$

so

$$\lim_{n\to\infty}[-\frac{1}{n}\log P_0\{Y_n < y_n'\}] = \lim_{n\to\infty}[-\frac{1}{n}\log \sqrt{y_n'}]$$

$$= \lim_{n\to\infty}[-\frac{n}{2}\log y_n']$$

$$= \nu^2 \ .$$

Now, if $\{P_n\}$ and $\{Q_n\}$ are two sequences of positive numbers such that $\lim_{n\to\infty}\frac{1}{n}\log P_n = \lim_{n\to\infty}\frac{1}{n}\log Q_n = k$,then, since $2\min(P_n,Q_n) \le P_n + Q_n \le 2\max(P_n,Q_n)$, it follows that $\frac{1}{n}\log(P_n+Q_n) \to k$ also.

Thus

$$-\frac{1}{n}\log P_0\{T_n^* < \sqrt{n}\nu\} = -\frac{1}{n}\log[P_0\{Y_n<y_n'\}+P_0\{Y_n>y_n''\}]$$

$$\to \nu^2 \ ,$$

so $f(\nu) = \nu^2$. Hence Theorem 1.1 gives $c(\mu)=2f(b(\mu))=2\cdot(\frac{\mu^2}{2}) = \mu^2$.

This conclusion is in accordance with the result of Chapter III which stated that the slope of $T_n^* = g_n(|\sqrt{n}\,\overline{X}|)$ cannot exceed the slope of $|\sqrt{n}\,\overline{X}|$. For this example $g_n(x) = |x|^{-1} e^{x^2/2}$, which is monotone increasing in $|\sqrt{n}\,\overline{X}|$ for values larger than one.

The importance of the next theorem to this thesis is that the proof is derived from the invariance property of exact slopes of Chapter III. The theorem is also given in (12). K is the Kullback-Liebler information function.

Theorem A.1: Let k be a one-to-one mapping of Θ onto itself, and let k^* be the mapping of Φ onto itself defined by $k^*(\phi) = k(\theta)$, where $\phi = g(\theta)$. Then $K(\theta, k(\theta)) = K^*(\phi, k^*(\phi))$.

Proof: Denote by $f(x, \theta)$ and $f^*(x, \phi)$ the density functions of x relative to Θ and Φ, respectively. Then $f(x,\theta) = f^*(x, g(\theta))$, where $\phi = g(\theta)$. Let $T(x) = \dfrac{f(x,\theta)}{f(x,k(\theta))}$ be the likelihood ratio statistic for testing H: $k(\theta)$ versus A: θ, let $T^*(x) = \dfrac{f^*(x,\phi)}{f^*(x,k^*(\theta))}$ be the likelihood ratio statistic for testing H^*: $k^*(\phi)$ versus A^*: ϕ. Then, by Theorem 1.2, $c(\theta) = 2K(\theta, k(\theta))$ and $c^*(\phi) = 2K^*(\phi, k^*(\phi))$. But for $\phi = k(\theta)$, $f(x,\theta) = f^*(x,\phi)$ gives $T(x) = T^*(x)$. Hence by invariance of exact slopes, $c(\theta) = c^*(\phi)$, so $K(\theta,k(\theta)) = K^*(\phi,k^*(\phi))$ for $\phi = g(\theta)$.

The last theorem shows a stronger superiority of the UI-P over FM than is given in (11).

Theorem A.2: Assume the situation of Example 4.1 for a fixed n, and for simplicity let $T_i = T_n^{(i)}$. For each α, $0 < \alpha < 1$, and each $\delta > 0$, the maximum type two error probability of the U-IP is smaller than the maximum type two error probability of FM, where the maximization is taken over the subset $\{(\theta_1,\theta_2): \theta_1 > \delta, \theta_2 > \delta\}$.

Proof: Let L_1 and L_2 be the levels attained by the tests T_1 and T_2, and let ℓ_α and $\tilde{\ell}_\alpha$ be defined by

$$P_0\{L_1 \cdot L_2 > \ell_\alpha\} = P_0\{L_1 > \tilde{\ell}_\alpha,\ L_2 > \tilde{\ell}_\alpha\} = 1-\alpha$$

Then ℓ_α and $\tilde{\ell}_\alpha$ are the distances from the coordinate axes to the points of intersection of the boundaries of the size α critical regions (in terms of levels) of FM and the UI-P, respectively, with the sides of the unit square. (See Figure 1). It is clear that $\ell_\alpha < \tilde{\ell}_\alpha$.

Now the type two error probabilities at the parameter point (θ_1, θ_2) for FM and the UI-P are given by $\beta(\alpha: \theta_1,\theta_2) = P_{\theta_1\theta_2}\{L_1 L_2 > \ell_\alpha\}$ and $\tilde{\beta}(\alpha; \theta_1,\theta_2) = P_{\theta_1\theta_2}\{L_1 > \tilde{\ell}_\alpha, L_2 > \tilde{\ell}_\alpha\}$. Thus

$$\beta(\alpha; \theta_1,\theta_2) = P_{\theta_1\theta_2}\{L_1 L_2 > \ell_\alpha\}$$

$$= \underset{\theta_1}{E}\,[P_{\theta_2}\{L_2 > \frac{\ell_\alpha}{L_1} \mid L_1\}]$$

$$\leq \underset{\theta_1}{E}\,[P_{-\infty}\{L_2 > \frac{\ell_\alpha}{L_1} \mid L_1\}] ,$$

since L_1 and L_2 are independent and $P_{\theta_2}\{L_2 > \ell\}$ is a montone decreasing function of θ_2. Now $P_{-\infty}\{L_2 > \ell\} = 1$ if $\ell < 1$ and $P_{-\infty}\{L_2 > \ell\} = 0$ if $\ell \geq 1$. Hence $\underset{\theta_1}{E}\,[P_{-\infty}\{L_2 > \frac{\ell_\alpha}{L_1} \mid L_1\}] = P_{\theta_1}\{L_1 > \ell_\alpha\}$, so

$$\sup_{\theta_i \geq \delta} \beta(\alpha; \theta_1,\theta_2) = P_\delta\{L_1 > \ell_\alpha\} .$$

Also

$$\tilde{\beta}(\alpha; \theta_1,\theta_2) = P_{\theta_1,\theta_2}\{L_1 > \tilde{\ell}_\alpha, L_2 > \tilde{\ell}_\alpha\}$$

$$= P_{\theta_1}\{L_1 > \tilde{\ell}_\alpha\}\, P_{\theta_2}\{L_2 > \tilde{\ell}_\alpha\}$$

$$\leq P_{\theta_1}\{L_1 > \tilde{\ell}_\alpha\} \; P_{-\infty}\{L_2 > \tilde{\ell}_\alpha\}$$

$$= P_{\theta_1}\{L_1 > \tilde{\ell}_\alpha\} \; .$$

Therefore

$$\sup_{\theta_i \geq \delta} \tilde{\beta}(\alpha;\; \theta_1, \theta_2) = P_\delta\{L_1 > \tilde{\ell}_\alpha\}$$

$$< P_\delta\{L_1 > \ell_\alpha\}$$

$$= \sup_{\theta_i \geq \delta} \beta(\alpha;\; \theta_1,\; \theta_2) \; ,$$

which is the desired result.

VITA

Ramon C. Littell

Candidate for the Degree of

Doctor of Philosophy

Thesis: PROPERTIES OF BAHADUR EFFICIENCY

Major Field: Statistics

Biographical:

Personal Data: Born in Rolla, Kansas, November 18, 1942, the son of Weldon H. and Laura Drew Littell.

Education: Attended elementary and high school in Rolla, Kansas; graduated from Rolla High School, 1960; received the Bachelor of Science degree from Kansas State Teachers College with a major in mathematics, May, 1964; received the Master of Science degree from Oklahoma State University, with a major in mathematics May, 1966; completed requirements for the Doctor of Philosophy degree in May, 1970, at Oklahoma State University.

Professional Experience: Graduate Assistant in the Department of Mathematics and Statistics, Oklahoma State University, 1964-1970.

Professional Organizations: American Statistical Association